THE UNIVERSITY
of EDINBURGH

Effects of external fluctuations on the stochastic dynamics of gene regulatory networks

Abhishek Gupta

Doctor of Philosophy
University of Edinburgh
2019

DECLARATION

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Abhishek Gupta
27 August 2019

ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof Ramon Grima, for giving me this opportunity, for all his support and patience during this period. I am thankful to my thesis committee members Prof Peter Swain and Dr Kostas Zygalakis, for sparing their time for fruitful meetings and suggestions. Once again, thanks to Prof Swain, for letting me spend a short span in his lab and learn about single cell microfluidic experiments. I am also thankful to Dr Shev MacNamara from University of Technology Syndney, for sharing his ideas and sparing his time.

I am thankful to Edinburgh Compute and Data Facility at University of Edinburgh, for providing high performance computing facility which saved lot of time for me. I would also like to thank CH Waddington Building administration and cleaning staff for providing a hygienic and clean work space, along with well organized everyday facilities.

I am thankful to all the past and present members of the Grima group. They are all amazing scientist and I wish them all the very best.

I am grateful to my family for standing by me, even when I couldn't perform and achieve anything significant.

Last but not the least, I want to express my gratitude toward my girlfriend Yixi Chen, who held my hand when I was falling behind, who shared her success with me and consoled me during all my failures. Without her, I don't know how I would have survived past couple of years. In times when I couldn't find anything positive around me, when I still don't hold any visible accolade, she is the only one I have, to speak to and to share my moments of relief and comfort. I am indebted to you for being around, for pushing me out of bed everyday and motivating me to stay afloat.

Doing a PhD, showed me what I could be at my worse. I just hope things get a bit better from here.

ABSTRACT

In systems biology, intracellular gene expression networks are modelled as mass balance chemical reactions. These schematic pathways are often simplified and miss significant details such as intermediate templates or enzymatic components, essentially needed in vivo. In our analysis, we explore further improvements in reach of these parsimonious models with inclusion of intermediate species as external components. Using stochastic simulations and complementary analytical methods, we compare mesoscopic behaviour of stand-alone original models and their interactions with external components. Implementing different examples of gene expression mechanisms, we demonstrate importance of upstream transcriptional noise as well as downstream fluctuations in protein degradations which are often ignored. We use transition in bimodal distributions as a germane tool to demonstrate these effects. Another pertinent example is our study on amplification of correlation resonance due to upstream transcription in a two protein translation network with proteolytic enzyme sharing. In conclusion, in this thesis, we try to bolster significance of underlying mechanistic details in modelling intracellular processes as well as emphasize pro and cons of different analytical approximation tools used in this study.

LAY SUMMARY

Mathematical models are an essential part of an iterative cycle of systems biology, where we transform experimental data into data-driven or mechanistic models and extract predictions to design new experiments. In this work, we have studied a few such models, which represent mathematical transformation of different intracellular mechanisms associated with central dogma and its regulation. Generally, numerical simulations are enforced to generate synthetic data by using brute force algorithms, to characterize and extract useful predictions from these models. In the context of intracellular biology, dynamics of various biomolecules is considered noisy due to 1) the randomness of interactions between them, and 2) a small number of molecular populations. Hence, Stochastic Simulation Algorithm (SSA) has been a reliable algorithm for years. But with the advancement of experimental methods, such as high throughput and high resolution techniques, discovery of new biomolecular interactions has further increased the complexity of these mathematical models. Hence, using simulations alone becomes computationally expensive and time-consuming. To resolve this, exact or approximate theoretical analysis is used to either substitute or compliment numerical simulations. In this regard, various analytical methods drawn from statistical and theoretical physics found numerical applications. For most of the time, performance measure for these exact or approximate analytical methods is that, how much their predictions agree with the ones made by numerical simulations.

While studying stochastic behaviour of different gene expression mechanisms, a mere schema of biological processes is used to construct mathematical models. The rationale behind making these simplifications is for the feasibility of exact theoretical analysis. Therefore, these models lack intricate details about intracellular interactions, such as intermediate templates and enzymatic machinery. Hence, very often, prediction reach of these models remains limited.

In our work, we tackle these limitations by exploring changes in stochastic behaviour of these parsimonious models due to interactions with a few dynamic intermediate species. While these dynamic intermediates are an essential part of biological systems, we construed these as sources of external fluctuations to our simplified models. Examples of these intermediates can be mRNA, ribosomes, proteolytic enzymes, etc.

With this increase in model complexity, most of the exact analytical approaches, which were derived for simplified models, don't work any more. Hence, we switch to various analytically approximate methods, which function seemingly well in parameter regimes, where inherent approximations hold true. With the use of these approximate methods, we studied a few noise driven characteristics of different gene expression mechanisms. The first one is bimodality, which refers to having two distinct peaks in a probability distribution. We implemented an approximate theoretical framework to study how different external fluctuations influence the bimodality exhibited during protein translation, in different mechanisms. The second one is Correlation Resonance, which refers to an extremum in correlation between two distinct proteins that share a common degradation enzyme. Our simulations and approximate analytical framework reveal a few interesting mechanistic aspects of Correlation Resonance, due to interactions of upstream external fluctuations with these mechanisms.

Overall, with our results, we signify the importance of approximation analytical methods, as well as the importance of intermediate template and catalytic molecules in intracellular biology.

LIST OF ABBREVIATIONS

RE	Rate Equation
SSA	Stochastic Simulation Algorithm
CME	Chemical Master Equation
ODE	Ordinary Differential Equation
CFPE	Chemical Fokker-Planck Equation
CLE	Chemical Langevin Equation
PDE	Partial Differential Equation
SDE	Stochastic Differential Equation
LNA	Linear Noise Approximation
FPE	Fokker-Planck Equation
SSE	System Size Equation
CDF	Cumulative Distribution Function
OU	Ornstein-Uhlenbeck
CC	Correlation Coefficient

CONTENTS

Chapter 1

INTRODUCTION

Stochasticity is an inherent property of all dynamic changes in nature. Though, due to a large number of interacting individuals, this probabilistic nature is often neglected and leads to a deterministic purview of macroscopic world [1]. In contrast, intracellular biology constitutes processes where inherent molecular fluctuations play decisive role due to a finite number of reacting individual species, for example, in context of cell biology, the number of protein molecules ranges in order of 10 to 10^4, similarly mRNA abundance varies from a single transcript to order of 10 [2,3]. Hence, in layman, cell biologists often deal with histograms or more specifically, probability distributions, means, variances and covariances of different dynamic species, instead of deterministic trajectories. Thus, we need to look beyond deterministic rate equations (RE) which fall short in these circumstances.

The traditional approach of investigating stochasticity in biochemical pathways is using Gillespie's Stochastic Simulation Algorithm (SSA) [4–6]. SSA is a Monte Carlo method which allows to simulate statistically exact sample trajectories. These trajectories can be further used for characterising features which emerge solely due to probabilistic nature of interactions.

While SSA is considered a stepping stone in quantifying effect of noise in cell and molecular biology, it requires extensive computational resources. This leads to difficulties in generating sample trajectories large enough to infer statistical properties with sufficient confidence [6], especially for pathways with higher number of species and those including reactions with slower time-scales.

To deal with this restriction, focus shifts to analytical methods which can provide an accurate estimation of stochastic behaviour of biochemical pathways. Chemical Master Equation (CME) is the first step in this direction [7]. CME is nothing but a form of Chapman-Kolmogorov equation for discrete Markov jump processes [8]. It represents a set of homogeneous linear first-order ordinary differential equations, capturing probabilities of reaction system being in any possible combination of molecules numbers of reacting species.

Even with a limited number of reaction molecules, most of biochemical pathways represent open systems where the total number of molecules is not fixed. Hence, in its exact form CME corresponds to a infinite set of Ordinary Differential Equations (ODEs) which is mostly dealt through numerical simulations and state space truncation [9]. In cases where multiple species are present, this analysis becomes almost infeasible due to huge possible combinations of number of molecules. Alternatively, focus switches to deriving exact or approximate solutions of CME. Exact time dependent transient solutions of CME are rare and mostly limited to relatively simpler biochemical systems, reaction networks of which consists of first order reactions [10] or single species single-step reactions [11]. To derive these solutions, a common approach is to transform CME using generating function. Stationary state solutions for biochemical pathways, even for those with multiple species and higher order chemical reaction is possible, if

detailed balance conditions are met [12], along with a few other specific cases [13–15]. Even with this respite, CME is generally considered analytically intractable and provides motivation for developing approximate solutions.

There are three major approximation approaches which set platform to deal with unmanageable CME. The first one is Chemical Fokker-Planck Equation/Chemical Langevin Equations (CFPE/CLE) which are derived using Taylor series expansion of CME upto second order [16, 17]. CFPE represents Fokker-Planck Partial Differential Equation (PDE) with variable drift and diffusion coefficients which approximate discrete Markov jump processes from CME to continuous diffusion processes [18]. While CFPE provides a convenient approximation of CME, its applicability as an analytical tool limits to one or two variable models or to models which can be reduced to similar dimensions. In contrast, CLE proves its utility as a relatively easier numerical analysis tool. Corresponding to a CFPE, CLE represents a set of stochastic differential equations (SDEs) [18, 19]. SDE benefits from a simple formulation as a combination of a drift and a diffusion term, which equal to RE and square roots of different reaction rates respectively. Due to their simple formulation and relatively faster simulations in comparison to discrete SSA trajectories, CLE finds wide popularity among computational biologists as well as in molecular and population epidemiology [20–22].

Second, in the row of approximation methods is Linear Noise Approximation (LNA) [23, 24]. In simple words, LNA depicts a small noise approximation due to sufficiently large number of molecules in a reaction system where trajectories fluctuate around its deterministic path [25, 26]. Fundamental principle behind LNA is the assumption that fluctuations in system state are inversely proportional to square root of system size Ω,

which from a statistician's purview parallels with central limit theorem. LNA comes from an expansion of CME around Ω and collecting coefficients of $\Omega^{-\frac{1}{2}}$ and Ω^0 respectively. Comparing coefficients of $\Omega^{-\frac{1}{2}}$ on both sides of CME gives rise to the same deterministic REs, which we can formulate using macroscopic mass action kinetics. Coefficients of Ω^0 lead to a linear Fokker-Planck equation (FPE) with constant drift and diffusion coefficients which with further transformation gives rise to a set of ODEs, capturing evolution of the covariance matrix. This final equation in stationary state conditions transforms to a continuous Lyapunov equation. In summary, LNA fits a Normal distribution around deterministic trajectories of the system and with increase in system size Ω, the agreement between system statistics calculated using SSA and LNA increases. LNA is one of the major tools which can be easily extended to complex networks with large number of constituents, hence it remains widely popular tool to characterize stochasticity and parameter inference in intracellular as well as population biology [27–31].

Another method (or set of methods) which helps in characterization of statistical properties of biochemical pathways is the moment closure approximation method [32–34]. Different moment closure approximations focus on estimation of stochastic properties (most importantly mean, variance and covariance) using different statistical moments, without much emphasis on inherent probability distribution. For reaction networks including only first order reactions, this approach renders exact moments. However, for reaction networks including reactions of second or higher orders, transforming CME to moment equations leads to an infinite open set of equations due to dependence of lower order moments on higher order moments. Hence, different approximations are proposed to close this open set of equations. Intuitively, limitations of different approximations arise in

form of discrepancies between analytical and simulation results which results from inherent information loss caused by approximation itself [35,36].

While analysing effects of inherent random nature of biochemical transformations, aforementioned methods are often supplemented with complementary transformations in the method itself as well as in the models. The transformations either make application of these analysis methods feasible when infeasible in their original form, or improve their performance. Some of these auxiliary techniques involve model reduction using time scale separation [37], model reduction based on difference in molecular abundance of different species [38] and variations in simulation algorithms such as $\tau-$leaping [39,40] to accelerate SSA simulations.

Development and applications of both simulation and analytical methods to quantify noise in biology have acted as the cog which was missing in the iterative cycle of systems biology [41]. Stochastic analysis of biochemical processes has revealed details which were previously ignored in macroscopic averaging, and the mechanistic reasoning behind them [42–47]. On all scales of biology, starting from central dogma to single cells as well as much beyond to microbial and mammalian populations, stochasticity has directed the cellular decision making, guided the phenotypic distributions and eventually shaped their evolutionary traits [1,48,49].

The first step in this elaborating quantitative analysis is to transform biochemical pathways into chemical reaction networks, from which we can write simple REs following mass action kinetics, as well as evaluate their statistical properties in context of stochasticity in biology. Most of these reaction networks and corresponding mathematical models are based on principles of parsimony. Preference to minimal models and keeping number of species in reaction networks as small as possible is always context

dependent. This preference is also based on the possibilities to derive an analytical solution for a simpler/smaller network with relative ease compared to a complicated/extended network [50]. A simple example of it could be understood using translation of proteins. If we are interested in knowing how upstream gene and mRNA stochastic dynamics influence protein distribution, we would be keen on reaction network as following, which represents translation from a single gene with constitutive promoter,

$$\begin{aligned} Gene &\xrightarrow{k_1} mRNA + Gene \\ mRNA &\xrightarrow{r_1} \Phi \\ mRNA &\xrightarrow{k_2} Protein + mRNA \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{1.1}$$

and if we are interested in knowing how various feedback mechanisms or downstream interactions shape protein distribution, we would focus on reaction network which would look something like this,

$$\begin{aligned} Gene &\xrightarrow{k_1} Protein + Gene \\ Gene^* &\xrightarrow{k_4} Protein + Gene^* \\ Protein &\xrightarrow{r_3} \Phi \\ Gene + 2\,Protein &\underset{k_3}{\overset{k_2}{\rightleftharpoons}} Gene^* \end{aligned} \tag{1.2}$$

which represents a non-linear self-regulatory feedback [15]. Following our discussion on parsimony and context dependence, it is clear that in mechanism from Eq. 1.2, focus is on understanding effects of feedback mechanisms, instead of transcription. Hence, assuming time scale separation [51] between dynamics of mRNA and dynamics of protein, missing mRNA in mechanism shown in Eq. 1.2 is justified. Same applies to

other intricate enzymatic and structural components of cellular machinery [52]. As expected, stochastic analysis using both simulation and analytical tools can provide ample information on effects of downstream feedback. However, what if assumption of time scale separation between transcription and translation (due to higher stability of mRNA [53, 54]) does not remain valid any more! In that case, mathematical model based on mechanism from Eq. 1.2 will fall short. It would be necessary to expand this mechanism by including mRNA transcription. Although it would still be possible to perform simulations using SSA (though simulations for this extended model would require more computational resources), exact analytical approach followed to derive analytical expression for distribution will no longer remain applicable.

In simple terminology, in context of mechanism shown in Eq. 1.2, we can approach stochastic dynamics of transcription as a source of ***external fluctuations*** which are extraneous and act in addition to inherent molecular noise of innate components of the mechanism [45, 55, 56]. In this thesis, we focus on demonstrate such interactions between a few unique biological mechanisms and different sources of external fluctuations, along with studying changes in stochastic behaviour of original mechanisms due to these interactions. With respect to protein translation, we have focused on both upstream and downstream external fluctuations. For example, transcription is upstream and fluctuations due to stochastic kinetics of proteolytic enzyme are downstream.

We start with chapter 2, building up mathematical foundation of simulation and analytical approximation methods. Along with SSA, CFPE/CLE and LNA, we also touched upon another specific approximation of FPE known as Unified Coloured Noise Approximation (UCNA) [57]. It provides an approximate transformation of single variable SDE incorporating coloured

noise (exponentially correlated with finite correlation time-scale) to the traditional white noise driven SDE. In chapter 3, we discuss effects of one such external fluctuations through different reaction channels in stochastic bimodality exhibited by Schlögl model [58]. It is one of the simplest, single species reaction networks, following mass action kinetics which exhibits macroscopic bistability and stochatic bimodality. As bimodality is prevalent in various aspects of cell and synthetic biology [59, 60], we choose it as a demonstration tool for these transitions led by external fluctuations. We implement a simple approximation to transform discrete external fluctuations into continuous diffusion process which represents, finite correlated coloured noise. With this transformation, a two-variable model can be reduced to a one-variable model, which brings UCNA into consideration. In chapter 4, we apply similar approach to model of gene expression and demonstrate applicability of UCNA as an approximate analytical tool to capture variations in protein translation through different gene regulatory mechanisms. In chapter 5, we explore another intriguing phenomena, known as Correlation Resonance [61, 62] and how it is affected by upstream transcription. In a translation network expressing two different proteins, Correlation Resonance represents an extremum in stochastic correlations, due to sharing of downstream proteolytic enzyme. Hence, with inclusion of transcriptional noise, we further explore different strategies of stochastic control in synthetic biology. Due to complexity of network, we switch to LNA as an approximate analytical method and compare it with SSA results. In chapter 6, we conclude with a small summary and discussion.

Chapter 2

PRELIMINARIES

In this chapter, we will provide a brief overview of established mathematical methods which are implemented in our research. As discussed in chapter 1, noise is a prevalent characteristic of functioning biology at all scales. To further our insights about underlying mechanisms of gene expression and effects of inherent molecular stochasticity, as a first step, we review various modelling aspects of stochastic chemical kinetics.

2.1 Deterministic versus Stochastic Kinetics

Applying law of mass action to reaction schemes gives rise to a set of differential equations depicting temporal dynamics of concentrations. A simple example using reaction,

$$A + B \xrightarrow{k} C \tag{2.1}$$

Here, we can keep track of deterministic kinetics using differential Eq. 2.2, also known as Rate Equation (RE),

$$\frac{dC}{dt} = k[A][B], \tag{2.2}$$

where [A] and [B] denote concentration of species A and B respectively and are conceived as continuous variables. k is the macroscopic reaction rate constant. For the same reaction in Eq. 2.1, through different means, stochastic kinetics keeps track of probability that molecules of species A and B take specific values at any time. Reaction takes place randomly and population of participating species can be addressed as discrete random variables n_A and n_B. Similar to mass action, reaction probability depends upon macroscopic reaction rate k and species populations n_A, n_B along with several other system parameters [6].

Therefore, stochastic dynamics of chemical reactions at mesoscopic scales associate reactive species with a probability density function which helps us quantify the random nature of molecular species and reactions. As molecule count becomes large, mesoscopic description becomes similar to macroscopic since fluctuations become negligible [63]. From biological perspective, mesoscopic purview renders a suitable framework to deal with reactions in distinctive intracellular conditions including finite small cellular volume and small number of biomolecules [64].

2.2 Chemical Master Equation (CME)

First derived by D. McQuarrie [65] and later on further popularized by D. Gillespie [4–7], CME captures stochastic dynamics of chemical reactions as Markovian jump processes. It is a primary equation which delineates the transient evolution of reaction system probabilities for being in any particular state. Here, state or system state refers to any achievable distinct combination of molecule numbers of reacting chemical species.

Let's consider, N distinct molecular species X_1;....;X_N participating in M

distinct elementary reactions R_1;....;R_M in a well stirred reaction volume Ω as following,

$$\sum_{i=1}^{N} s_{ij} X_i \xrightarrow{k_j} \sum_{i=1}^{N} r_{ij} X_i \tag{2.3}$$

Here, s_{ij} and r_{ij} are stoichiometric coefficients for the ith species on reactant and product sides of the jth reaction respectively. k_j is the macroscopic rate coefficient of the jth reaction. We can define a system state vector as random variable $\boldsymbol{Y}(t) = [Y_1(t); ....; Y_N(t)]$. Let's say, $\boldsymbol{Y}(t) = \boldsymbol{n}$ is a realization of $\boldsymbol{Y}$ where $\boldsymbol{n} = [n_1; ....; n_N]$ and each n_i confers to number of molecules species X_i has in system at time t. In a very small time interval Δt, when only a single reaction could take place, let's say jth reaction occurs, thus changing system state from $\boldsymbol{n}$ to $(\boldsymbol{n} + \boldsymbol{\nu}_j)$. $\boldsymbol{\nu}_j = [(r_{1j} - s_{1j}); ....; (r_{Nj} - s_{Nj})]$ denotes change in the number of molecules of each species with occurrence of the jth reaction. $\boldsymbol{\nu}_j$ represents the jth column of the stoichiometric matrix $\boldsymbol{\nu}$. We can write probability of system being in state $\boldsymbol{n}$ at time $t + \Delta t$ as,

$$\begin{aligned} P(\boldsymbol{n}, t + \Delta t) &= P(\boldsymbol{n}, t) + \Delta t \sum_{j=1}^{M} w_{j,(\boldsymbol{n}-\boldsymbol{\nu}_j)\to\boldsymbol{n}} P((\boldsymbol{n} - \boldsymbol{\nu}_j), t) ... \\ &\quad ... - \Delta t \sum_{j=1}^{M} w_{j,\boldsymbol{n}\to(\boldsymbol{n}-\boldsymbol{\nu}_j)} P(\boldsymbol{n}, t), \end{aligned} \tag{2.4}$$

In above equation, on the left hand side $P(\boldsymbol{n}, t + \Delta t)$ is probability of reaction system being in state $\boldsymbol{n}$ at time $t + \Delta t$. On the right hand side, first term $P(\boldsymbol{n}, t)$ represents probability of system being in state $\boldsymbol{n}$ at time t, second term, $\Delta t \sum_{j=1}^{M} w_{j,(\boldsymbol{n}-\boldsymbol{\nu}_j)\to\boldsymbol{n}} P((\boldsymbol{n} - \boldsymbol{\nu}_j), t)$ represents cumulative probability of arrival in state $\boldsymbol{n}$ from all other plausible states $(\boldsymbol{n} - \boldsymbol{\nu}_j)$. Similarly, third term $\Delta t \sum_{j=1}^{M} w_{j,\boldsymbol{n}\to(\boldsymbol{n}-\boldsymbol{\nu}_j)} P(\boldsymbol{n}, t)$ is cumulative probability of departure from state $\boldsymbol{n}$ to all other plausible states $(\boldsymbol{n} - \boldsymbol{\nu}_j)$. $w_{j,(\boldsymbol{n}-\boldsymbol{\nu}_j)\to\boldsymbol{n}} \Delta t$ and $w_{j,\boldsymbol{n}\to(\boldsymbol{n}-\boldsymbol{\nu}_j)} \Delta t$ are transition probabilities between different states, denoting probabilities of

occurrence of the jth reaction at system states $(\boldsymbol{n} - \boldsymbol{\nu}_j)$ and $\boldsymbol{n}$ in time interval Δt. For an infinitesimally small Δt where $\Delta t \to dt$, Eq. 2.4 can be rearranged as,

$$\frac{dP(\boldsymbol{n},t)}{dt} = \sum_{j=1}^{M} \left(w_{j,(\boldsymbol{n}-\boldsymbol{\nu}_j)} P((\boldsymbol{n}-\boldsymbol{\nu}_j),t) - w_{j,\boldsymbol{n}} P(\boldsymbol{n},t) \right) \tag{2.5}$$

Eq. 2.5 formulates the widely used generic form of CME [63]. In above equation, state transitions $(\boldsymbol{n} - \boldsymbol{\nu}_j) \to \boldsymbol{n}$ and $\boldsymbol{n} \to (\boldsymbol{n} - \boldsymbol{\nu}_j)$ have been abbreviated for conciseness. For a reaction system, $\boldsymbol{w_n} = [w_{1,\boldsymbol{n}}; w_{2,\boldsymbol{n}}; ...; w_{M,\boldsymbol{n}}]$ are the *mesoscopic propensities* [6, 63] of participating reactions. For elementary reactions following mass action kinetics, it can be written as,

$$w_{j,\boldsymbol{n}} \;=\; k_j \Omega \prod_{i=1}^{N} \frac{n_i!}{(n_i - s_{ij})! \Omega^{s_{ij}}}, \tag{2.6}$$

Table 2.1 shows propensities for various elementary reactions. For different reaction orders, Ω contributes to balance propensity dimensionally as macroscopic rate constants include $Volume^{(Order-1)}$ in its units. This dependency of reaction propensities on system size (volume) gives an intuitive idea about reaction probabilities. In accordance with the collision theory, inverse proportionality of reaction propensity to $\Omega^{Order-1}$ indicates smaller chance of a reaction that requires collision of a large number of molecules to collide together, thus making reactions with order three or more highly improbable.

Table 2.1: Elementary reactions with their corresponding propensity

Reaction	**Order**	**Propensity**	**Units of k**
$\phi \xrightarrow{k}$	Zero	$k\Omega$	$Volume^{-1}Time^{-1}$
$S \xrightarrow{k}$	First	kn	$Time^{-1}$
$S_1 + S_2 \xrightarrow{k}$	Second	$\frac{kn_1n_2}{\Omega}$	$VolumeTime^{-1}$
$2S \xrightarrow{k}$	Second	$\frac{n(n-1)}{\Omega}$	$VolumeTime^{-1}$
$S_1 + S_2 + S_3 \xrightarrow{k}$	Third	$\frac{kn_1n_2n_3}{\Omega^2}$	$Volume^2Time^{-1}$
$2S_1 + S_2 \xrightarrow{k}$	Third	$\frac{kn_1(n_1-1)n_2}{\Omega^2}$	$Volume^2Time^{-1}$
$3S \xrightarrow{k}$	Third	$\frac{kn(n-1)(n-2)}{\Omega^2}$	$Volume^2Time^{-1}$

2.2.1 Examples

For chemical reaction given in Eq. 2.1, we can write CME as,

$$\frac{dP(n_A,n_B,n_C,t)}{dt} = \frac{k}{\Omega}\Big((n_A+1)(n_B+1)P((n_A+1),(n_B+1),(n_C-1),t)... \\ ... - n_An_BP(n_A,n_B,n_C,t)\Big), \tag{2.7}$$

here, n_A, n_B, n_C denotes numbers of molecules of species A, B, C respectively. k is macroscopic rate constant and Ω is reaction volume. Hence, Eq. 2.7 comprises a set of differential equations where each equation captures evolution of system state probability for a unique combination of n_A, n_B, n_C. Reaction 2.1 represents a 1-step jump process in a closed system and follows conservation laws (i) $n_A + n_B + n_C = n_{0_A} + n_{0_B} + n_{0_C}$ and (ii) $n_A - n_B = n_{0_A} - n_{0_B}$. $n_{0_A}, n_{0_B}, n_{0_C}$ are numbers of molecules at $t = 0$, thus Eq. 2.7 can be reduced to a single variable equation. A systematic linear algebraic method that provides both transient and stationary state solutions is provided in [11].

Similarly, for a biochemical reaction system as shown below representing simple schematic of a 2-stage gene expression model comprising a single constitutive promoter that doesn't require any inducer, suppressor or signalling to start translation,

$$\begin{aligned} Gene &\xrightarrow{k_1} mRNA + Gene \\ mRNA &\xrightarrow{r_1} \Phi \\ mRNA &\xrightarrow{k_2} Protein + mRNA \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{2.8}$$

CME can be formulated as shown in Eq. 2.9 [14, 66]. Here, n_1 and n_2 denote the number of individual mRNA and protein molecules respectively.

$$\begin{aligned} \frac{dP(n_1,n_2,t)}{dt} &= k_1\Big(P((n_1-1),n_2,t) - P(n_1,n_2,t)\Big) + r_1\Big((n_1+1)P((n_1+1),n_2,t)... \\ &\quad ... - n_1P(n_1,n_2,t)\Big) + k_2\Big(n_1P(n_1,(n_2-1),t) - n_1P(n_1,n_2,t)\Big)... \\ &\quad ... + r_2\Big((n_2+1)P(n_1,(n_2+1),t) - n_2P(n_1,n_2,t)\Big), \end{aligned} \tag{2.9}$$

An approximated analytical solution of Eq. 2.9 is provided in [14] for protein degradation time scales, which are much larger than mRNA degradation time scales ($\frac{1}{r_2} \gg \frac{1}{r_1}$). As complexity of biochemical pathways increases with increase in the number of molecular interactions as well as non-linearity of these interactions, CME becomes mathematically intractable and remains solvable for a very few specific cases. In such conditions, CME can be further transformed into moment equations using moment generating function [66], which are exactly solvable for linear reaction networks (similar to Eq. 2.8) or can be further simplified using various moment closure

approximations [32–34]. Other analytical approximations include CFPE/CLE and SSE which are used to extract useful statistical properties of stochastic biochemical pathways.

2.3 Stochastic Simulation Algorithm (SSA)

Although numerical simulations of CME itself provide brute force alternative to analytical handling of CME, with possibility of either large or infinite number of system states, handling of such vast number of ODEs becomes computationally infeasible [9, 63]. In contrast, SSA provides a practical and easily implementable simulation method [4, 5]. Instead of simulating transient evolution of probability distribution, SSA provides a route to sample statistically correct trajectories of reacting species itself. To create a stationary state probability distribution, following the ergodic nature of Markov processes, we can either sample one really long trajectory or multiple small trajectories.

SSA is a Monte Carlo simulation method and is based on exponential distribution of reaction waiting times. It assumes, in an infinitesimally small time interval, only one reaction can take place, almost instantly and there will be an exponentially distributed waiting time τ, before a second such reaction event can take place. Intuitively, mean of waiting time between reaction events is inverse of sum of all reaction propensities as shown in Eq. 2.10. Thus, this distribution keeps on changing as reaction system progresses in time and attains different states.

$$\tau \sim exp\left(\sum_{j=1}^{M} w_{j,\boldsymbol{n}}\right) \tag{2.10}$$

In practice, inverse transform sampling is used to generate exponentially

distributed random numbers. For an exponential distribution with a cumulative distribution function (CDF),

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2.11}$$

where $\lambda = \frac{1}{\sum_{j=1}^{M} w_{j,\boldsymbol{n}}}$, we can obtain inverse by solving for $y = F_X(x)$, which gives,

$$x = -\frac{1}{\lambda} ln(1 - y), \tag{2.12}$$

Thus, if we pick a uniformly distributed random variable r_1 from $U(0,1)$, then random variable $-\frac{1}{\lambda} ln(1 - r_1)$ or $-\frac{1}{\lambda} ln(r_1)$ will have an exponential distribution.

After each waiting interval, one reaction is selected randomly, where chance of selection depends solely on reaction propensities. For a reaction system consisting of M reactions with propensities $[w_{1,\boldsymbol{n}}; w_{2,\boldsymbol{n}}; ...; w_{M,\boldsymbol{n}}]$, we can write normalize propensities as $[\frac{w_{1,\boldsymbol{n}}}{w_{\boldsymbol{n}o}}; \frac{w_{2,\boldsymbol{n}}}{w_{\boldsymbol{n}o}}; ...; \frac{w_{M,\boldsymbol{n}}}{w_{\boldsymbol{n}o}}]$, where $w_{\boldsymbol{n}o}$ is sum of propensities $\sum_{j=1}^{M} w_{j,\boldsymbol{n}}$. To select a reaction, first we pick another uniformly distributed random variable r_2 from $U(0,1)$. Secondly, after fixing reaction indices, we find the smallest reaction indices J which follows,

$$J = \underset{j}{\operatorname{argmin}} \sum_{j=1}^{M} \frac{w_{j,\boldsymbol{n}}}{w_{\boldsymbol{n}o}} > r_2, \tag{2.13}$$

Algorithms proceeds iteratively, (i) defining reaction propensities at present system state, (ii) sampling τ as mentioned in Eq. 2.10, (iii) executing chosen reaction according to Eq. 2.13 and updating system states (iii) advancing time by τ.

SSA finds wide applications in systems biology. It is still a primal tool for stochastic analysis of gene expression and relevant population dynamic models [67, 68]. However it limits capabilities of analysis due to required computational expenditure. It requires sampling of large number of individual trajectories to capture statistical properties with sufficient accuracy, if system fluctuations are large due to limited number of reacting molecules. Another bottleneck is the extremely small waiting time intervals between reaction events due to high number of molecules, which makes SSA extremely slow to reach truncation time.

2.4 Chemical Fokker-Planck Equation (CFPE) and Chemical Langevin Equation (CLE)

Analytical treatment of CME remains limited to either a few specific cases based on reaction network topology or to innovative approximations which provide further simplifications. Similarly, applications of moment equations are constricted due to coupling of different moment equations leading to an infinite series of equations. Hence, different approximation methods to close these open system of equations, influence the derived moments.

In contrast, CFPE and its complementary CLE allow approximation of discrete jump Markov processes to continuous diffusion Markov processes [18, 19]. For linear reaction systems, where reaction order is no more than first order, moments derived from CFPE/CLE match exactly with moments obtained using exactly solvable moment equations from CME. In theoretical analysis of reaction networks, the use of CFPE is mostly limited to cater an analytical stationary state distribution for single species models [59, 69], beyond that it becomes difficult to deal with, both

analytically and numerically. However, CLE is still a much more popular analysis tool because of explicit formulation and ease of simulations even for much larger systems [20–22], in comparison to CME and SSA.

CFPE was first derived by Kramers and Moyal [16, 17] based on a simple assumption of system state $\boldsymbol{n}$ being continuous real numbers instead of discrete integers. Let's consider, in a reaction system, the number of molecules of any individual species n_i is large enough so that for any reaction j, $\nu_{ij} \ll n_i$ or change in the number of molecules due to any reaction is much smaller than the total molecules of any individual species. In these conditions, expanding Eq. 2.5 around system state $\boldsymbol{n}$ using Taylor's series upto second order, yields the well known CFPE as,

$$\frac{dP(\boldsymbol{n},t)}{dt} = \sum_{j=1}^{M}\left(-\sum_{i=1}^{N}\nu_{ij}\frac{\partial}{\partial n_i}\left[w_{j,\boldsymbol{n}}P(\boldsymbol{n},t)\right] + \frac{1}{2}\sum_{i,k=1}^{N}\nu_{ij}\nu_{kj}\frac{\partial^2}{\partial n_i \partial n_k}\left[w_{j,\boldsymbol{n}}P(\boldsymbol{n},t)\right]\right) \tag{2.14}$$

Although we can simply write corresponding Langevin equation to above FPE, the actual derivation of CLE was presented independently, without invoking the Taylor's series truncation of CME [19]. Let's consider a time interval Δt, which follows:

(i) small enough that change in any reaction propensity $w_{j,\mathbf{Y}(t')}$ in time interval $[t, t+\Delta t]$ is insignificant,

$$w_{j,\mathbf{Y}(t')} \cong w_{j,\mathbf{Y}(t)=\boldsymbol{n}}, \qquad t' \in [t, t+\Delta t], \tag{2.15}$$

thus, we can assume number of occurrences of each reaction in interval $[t, t+\Delta t]$ behaves as an independent Poisson distributed random variable. In this condition, we can write system state at time $t+\Delta t$, as

random variable $\boldsymbol{Y}(t + \Delta t)$ which follows:

$$\boldsymbol{Y}(t + \Delta t) = \boldsymbol{n} + \sum_{j=1}^{M} \boldsymbol{\nu}_j \mathcal{P}_j(w_{j,\boldsymbol{n}}, \Delta t), \tag{2.16}$$

As, mentioned previously, $\boldsymbol{n}$ is a realization of system state $\mathbf{Y}(t)$ and $\boldsymbol{\nu}_j$ is the change in number of molecules with occurrence of the jth reaction. $\mathcal{P}_j(w_{j,\boldsymbol{n}}, \Delta t)$ is an independent Poisson random variable, representing number of occurrences of each reaction.

(ii) large enough, so that expectation value of occurrences of each reaction $\langle \mathcal{P}_{ij}(w_{j,\boldsymbol{n}}, \Delta t) \rangle$ is much larger than 1, hence $\mathcal{P}_{ij}(w_{j,\boldsymbol{n}}, \Delta t)$ can be approximated with statistically independent normal random variable $\mathcal{N}(w_{j,\boldsymbol{n}}\Delta t, w_{j,\boldsymbol{n}}\Delta t)$. Thus, Eq. 2.16 can be further transformed as,

$$\boldsymbol{Y}(t + \Delta t) = \boldsymbol{n} + \sum_{j=1}^{M} \boldsymbol{\nu}_j \mathcal{N}_j(w_{j,\boldsymbol{n}}\Delta t, w_{j,\boldsymbol{n}}\Delta t), \tag{2.17}$$

or using linear transformation of any Normal distribution to standard Normal distribution $\mathcal{N}_j(0, 1)$, we can write,

$$\boldsymbol{Y}(t + \Delta t) = \boldsymbol{Y} + \sum_{j=1}^{M} \boldsymbol{\nu}_j w_{j,\boldsymbol{Y}} \Delta t + \sum_{j=1}^{M} [w_{j,\boldsymbol{Y}} \Delta t]^{\frac{1}{2}} \boldsymbol{\nu}_j \mathcal{N}_j(0, 1), \tag{2.18}$$

Approximating discrete Poisson distributed random variables with continuous Normally distributed random variables provides approximate transformation of discrete Markov jump processes into continuous Markov diffusion processes, similar to the assumption made in Kramer-Moyal's derivation of CFPE. While conditions in (i) and (ii) seem mutually contradicting, but for reacting systems with sufficiently large molecule numbers, both conditions can be true simultaneously. In most cases reaction stoichiometry coefficients are no more than 2 or 3, hence the change in

molecule numbers due to any reaction ν_{ij} is always limited to a small integer. Sufficiently large number of molecules in a reaction volume ensures that even if any reaction occurs multiple times, change in molecule number remains small in comparison to total number of molecules. For an infinitesimally small Δt, when $\Delta t \to dt$, Eq. 2.18 can be represented in form of a Langevin equation incorporating White-Noise:

$$\frac{d\boldsymbol{Y}(t)}{dt} = \sum_{j=1}^{M} \boldsymbol{\nu}_j w_{j,\boldsymbol{Y}} + \sum_{j=1}^{M} w_{j,\boldsymbol{Y}}^{\frac{1}{2}} \boldsymbol{\nu}_j \Gamma_j \tag{2.19}$$

where, Γ_j is defined as $\Gamma_j(t) \equiv lim_{dt\to 0}\mathcal{N}_j(0, 1/dt)$, representing temporally uncorrelated, statically independent *Gaussian white noise*.

Thus, CLE is derived without evoking complementarity with Kramer-Moyal's FPE. It avoids the initial baffling assumption of molecule numbers as continuous real variables, instead, it provides a theoretical account behind discrete to continuous transformation of molecule numbers.

2.5 System Size Expansion (SSE): A Linear Noise Approximation (LNA)

Another popular method to approximate CME is known as System Size Expansion (SSE) [23, 24], where CME is expanded around system size Ω. In this expansion, terms of order Ω^0 yield a FPE with a linear drift term and a constant diffusion term. This new FPE approximates distribution of inherent fluctuations based on reaction network stoichiometry and propensities. Hence, upon neglecting further higher order terms $\mathcal{O}(\Omega^{-\frac{1}{2}})$, this analytical approach for approximating probability distributions of reactions networks,

can be addressed as a Linear Noise Approximation (LNA) [25,26].

Basic premise of SSE comes from separation of system states $\boldsymbol{n}$ into deterministic concentrations $\boldsymbol{\phi}$ and fluctuations $\boldsymbol{\varepsilon}$ of order $\Omega^{1/2}$. Here, $\boldsymbol{\phi} = [\phi_1; ...; \phi_N]$ and $\boldsymbol{\varepsilon} = [\varepsilon_1; ...; \varepsilon_N]$, are vectors comprising deterministic concentrations and fluctuations in molecules of individual species. Thus, we can write ansatz,

$$\boldsymbol{n} = \Omega\boldsymbol{\phi} + \Omega^{1/2}\boldsymbol{\varepsilon}, \tag{2.20}$$

In Eq. 2.20, $\Omega\boldsymbol{\phi}$ and $\Omega^{1/2}\boldsymbol{\varepsilon}$ can be regarded as deterministic macroscopic and stochastic mesoscopic contributions to the system state $\boldsymbol{n}$. Thus, as $\Omega \to \infty$ leads to $\frac{\boldsymbol{n}}{\Omega} \to \boldsymbol{\phi}$.

As a first step, we can define a step-operator as following,

$$\mathbf{E}_i^{-\nu_{ij}} f(..., n_i, ...) = f(..., (n_i - \nu_{ij}), ...), \tag{2.21}$$

Secondly, with the help of operator defined in Eq. 2.21, we can transform CME described in Eq. 2.5 as,

$$\frac{dP(\boldsymbol{n}, t)}{dt} = \sum_{j=1}^{M} \left(\prod_{i=1}^{N} \mathbf{E}_i^{-\nu_{ij}} - 1 \right) w_{j,\boldsymbol{n}} P(\boldsymbol{n}, t), \tag{2.22}$$

Now, let's assume probability distribution that captures fluctuations $\boldsymbol{\varepsilon}$ is $\Pi(\boldsymbol{\varepsilon}, t)$. Using change of variable, we can transform probability distribution $P(\boldsymbol{n}, t)$ as,

$$\Pi(\boldsymbol{\varepsilon}, t) = \left| \frac{\partial \boldsymbol{n}}{\partial \boldsymbol{\varepsilon}} \right| P(\boldsymbol{n}, t) = \Omega^{\frac{N}{2}} P(\boldsymbol{n}, t), \tag{2.23}$$

Eq. 2.23 represents probability of change in system state by $d\boldsymbol{n}$ is equal to probability of change in system fluctuations by $d\boldsymbol{\varepsilon}$. Applying chain rule, we

can write,

$$\frac{dP(\boldsymbol{n},t)}{dt} = \Omega^{-N/2}\left(\frac{\partial \Pi(\boldsymbol{\varepsilon},t)}{\partial t} - \Omega^{1/2}\sum_{i=1}^{N}\frac{d\phi_i}{dt}\frac{\partial \Pi(\boldsymbol{\varepsilon},t)}{\partial \varepsilon_i}\right), \tag{2.24}$$

Using step-operator in Eq. 2.21, we can write,

$$\prod_{i=1}^{N}\mathbf{E}_i^{-\nu_j} f(\boldsymbol{n}) = f(\boldsymbol{n}-\boldsymbol{\nu}_j), \tag{2.25}$$

which with the help of ansatz in Eq. 2.20 and Taylor's series expansion upto second order, can be written as,

$$\prod_{i=1}^{N}\mathbf{E}_i^{\nu_{ij}} f(\Omega\boldsymbol{\phi}+\Omega^{1/2}\boldsymbol{\varepsilon}) \approx f(\Omega\boldsymbol{\phi}+\Omega^{1/2}\boldsymbol{\varepsilon}) - \sum_{i=1}^{N}\frac{\nu_{ij}}{\sqrt{\Omega}}\frac{\partial f(\Omega\boldsymbol{\phi}+\Omega^{1/2}\boldsymbol{\varepsilon})}{\partial \varepsilon_i} + \sum_{i,k=1}^{N}\frac{\nu_{ij}\nu_{kj}}{2\Omega}\frac{\partial^2 f(\Omega\boldsymbol{\phi}+\Omega^{1/2}\boldsymbol{\varepsilon})}{\partial \varepsilon_i \varepsilon_k}, \tag{2.26}$$

Thus, we can write,

$$\prod_{i=1}^{N}\mathbf{E}_i^{\nu_{ij}} - 1 \approx -\sum_{i=1}^{N}\frac{\nu_{ij}}{\sqrt{\Omega}}\frac{\partial}{\partial \varepsilon_i} + \sum_{i,k=1}^{N}\frac{\nu_{ij}\nu_{kj}}{2\Omega}\frac{\partial^2}{\partial \varepsilon_i \varepsilon_k}, \tag{2.27}$$

While expanding propensities, first using Eq. 2.6, we write,

$$w_{j,\boldsymbol{n}} = \Omega w_{j,\frac{\boldsymbol{n}}{\Omega}}, \tag{2.28}$$

where, $w_{j,\frac{\boldsymbol{n}}{\Omega}}$ are transformed propensities as function of $\frac{\boldsymbol{n}}{\Omega}$. In Eq. 2.20, we can see that $\Omega \to \infty$ leads to $\frac{\boldsymbol{n}}{\Omega} \to \boldsymbol{\phi}$. Hence, we can expand reaction propensities $w_{j,\boldsymbol{n}}$ upto first order of Taylor's series expansion while $\Omega \to \infty$ as,

$$w_{j,\boldsymbol{n}} = \Omega w_{j,\frac{\boldsymbol{n}}{\Omega}} \approx \Omega\left(w_{j,\boldsymbol{\phi}} + \frac{1}{\sqrt{\Omega}}\sum_{l=1}^{N}\varepsilon_l\frac{\partial w_{j,\phi}}{\partial \phi_l}\right), \tag{2.29}$$

Using equations 2.23, 2.24, 2.27 and 2.29, we can expand CME in Eq. 2.22 as

following,

$$\left(\frac{\partial\Pi(\boldsymbol{\varepsilon},t)}{\partial t}-\Omega^{1/2}\sum_{i=1}^{N}\frac{d\phi_i}{dt}\frac{\partial\Pi(\boldsymbol{\varepsilon},t)}{\partial\varepsilon_i}\right)=\sum_{j=1}^{M}\left(-\sum_{i=1}^{N}\frac{\nu_{ij}}{\sqrt{\Omega}}\frac{\partial}{\partial\varepsilon_i}+\sum_{i,k=1}^{N}\frac{\nu_{ij}\nu_{kj}}{2\Omega}\frac{\partial^2}{\partial\varepsilon_i\varepsilon_k}\right)\ldots$$
$$\ldots\Omega\left(w_{j,\boldsymbol{\phi}}+\frac{1}{\sqrt{\Omega}}\sum_{l=1}^{N}\varepsilon_l\frac{\partial w_{j,\phi}}{\partial\phi_l}\right)\Pi(\boldsymbol{\varepsilon},t), \tag{2.30}$$

From Eq. 2.30, collecting coefficients of $\Omega^{1/2}$ on both sides returns macroscopic REs,

$$\frac{d\phi_i}{dt}=\sum_{j=1}^{M}\nu_{ij}w_{j,\phi}, \tag{2.31}$$

or in matrix form,

$$\frac{d\boldsymbol{\phi}}{dt}=\boldsymbol{\nu}.\boldsymbol{w}_{\boldsymbol{\phi}}, \tag{2.32}$$

Similarly, collecting coefficients of Ω^0 on both sides returns a FPE,

$$\frac{\partial\Pi(\boldsymbol{\varepsilon},t)}{\partial t}=\sum_{j=1}^{M}\left(-\sum_{i,l=1}^{N}\nu_{ij}\frac{\partial w_{j,\boldsymbol{\phi}}}{\partial\varepsilon_l}\frac{\partial}{\partial\varepsilon_i}\varepsilon_l+\frac{1}{2}\sum_{i,k=1}^{N}\nu_{ij}\nu_{kj}w_{j,\boldsymbol{\phi}}\frac{\partial^2}{\partial\varepsilon_i\varepsilon_k}\right)\Pi(\boldsymbol{\varepsilon},t), \tag{2.33}$$

or in matrix form,

$$\frac{\partial\Pi(\boldsymbol{\varepsilon},t)}{\partial t}=-\left(\boldsymbol{J}\nabla\cdot(\boldsymbol{\varepsilon}\Pi(\boldsymbol{\varepsilon},t))-\frac{1}{2}\boldsymbol{D}\nabla^2\Pi(\boldsymbol{\varepsilon},t)\right), \tag{2.34}$$

where, $J_{il}=\sum_{j=1}^{M}\nu_{ij}\frac{\partial w_{j,\boldsymbol{\phi}}}{\partial\varepsilon_l}$ are elements of Jacobian matrix derived from RE 2.31 and $D_{ik}=\sum_{j=1}^{M}\nu_{ij}\nu_{kj}w_{j,\boldsymbol{\phi}}$ are elements of Diffusion matrix. ∇ and ∇^2 are Divergence and Laplacian operators respectively.

As mentioned previously and evident in Eq. 2.34, LNA provides a FPE consisting of a drift term as linear function of system fluctuations and a

constant diffusion term evaluated at macroscopic concentrations. With these characteristics, Eq. 2.34 denotes an Ornstein–Uhlenbeck (OU) process which exhibits a Normal distribution both in transient and stationary states and has a feasible exact solution, written as,

$$\Pi(\boldsymbol{\varepsilon}, t) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C}|}} e^{-\frac{1}{2}\boldsymbol{\varepsilon}^T \boldsymbol{C}^{-1}(t)\boldsymbol{\varepsilon}}, \tag{2.35}$$

where $\boldsymbol{C}$ is time dependent covariance matrix, which can be derived from FPE shown in Eq. 2.34,

$$\frac{d\boldsymbol{C}}{dt} = \boldsymbol{J} \cdot \boldsymbol{C} + \boldsymbol{C} \cdot \boldsymbol{J} + \boldsymbol{D}, \tag{2.36}$$

with initial condition $\boldsymbol{C}(t = 0) = 0$. $\boldsymbol{C}$ captures variance and covariance of system fluctuations $\boldsymbol{\varepsilon}$ and can be transformed to variance and covariance of system state $\boldsymbol{n}$ by scaling with Ω. Eq. 2.36 is also known as Lyapunov equation [27, 70]. Similarly, using equation 2.34, expectation of ε, $\langle \boldsymbol{\varepsilon} \rangle$ can be written as,

$$\frac{d\langle \boldsymbol{\varepsilon} \rangle}{dt} = \boldsymbol{J} \cdot \langle \boldsymbol{\varepsilon} \rangle, \tag{2.37}$$

Thus, initial condition $\langle \boldsymbol{\varepsilon} \rangle (t = 0) = 0$ leads to $\langle \boldsymbol{\varepsilon} \rangle = 0$ for all time, implying system state expectation $\langle \boldsymbol{n} \rangle$ equals to $\boldsymbol{\phi}\Omega$.

At stationary state, Eq. 2.36 further simplifies to,

$$\boxed{\boldsymbol{J}_S \cdot \boldsymbol{C}_S + \boldsymbol{C}_S \cdot \boldsymbol{J}_S + \boldsymbol{D}_S = 0} \tag{2.38}$$

where, $\boldsymbol{C}_S$ is stationary state covariance matrix with $\boldsymbol{J}_S$ and $\boldsymbol{D}_S$ calculated at stationary state deterministic concentration $\boldsymbol{\phi}_S$. Equation 2.38 represents stationary state fluctuation-dissipation relation between drift and diffusion

terms of FPE in Eq. 2.34.

While LNA fails to capture deviations in system state expectations $\langle \boldsymbol{n} \rangle$, SSE with higher order terms including coefficients of $\Omega^{-1/2}$ successfully resolves this limitation of LNA [71].

2.6 Unified Coloured Noise Approximation (UCNA)

So far we have seen that FPE plays a major role in analytical approximation of stochasticity in biochemical pathways. Applications of non-linear CFPEs are mostly limited to single variable models. In contrast, a further simplified linear FPE derived using SSE and LNA with some compromise on accuracy, finds wider applications. On this similar note, UCNA advances applications of single variable FPEs for non-Markovian coloured noise processes with finite correlation time [57,72]. As we saw, CME encapsulates dynamics of chemical reactions in the form of memoryless Markovian jump processes, which can be further approximated in form of CLE driven by δ-correlated White-noise. Hence, UCNA proves its utility by providing an approximate stationary state solution for single variable Langevin equation with Gaussian coloured noise. In this section, we provided a detailed derivation of UCNA for a Langevin equation driven by both multiplicative Gaussian White and Coloured noise terms, along with its corresponding approximate FPE and stationary state distribution.

To begin with let's consider a stochastic differential equation,

$$\dot{x} \;=\; h(x) + g_1(x)\epsilon(t) + g_2(x)\Gamma(t), \tag{2.39}$$

where ϵ is Gaussian coloured noise, following OU process,

$$\dot{\epsilon} = -\frac{1}{\tau}\epsilon + \frac{1}{\tau}\Gamma_1(t), \tag{2.40}$$

and $\Gamma(t)$ and $\Gamma_1(t)$ are Gaussian white noises,

$$\begin{aligned} \langle \Gamma(t) \rangle &= \langle \Gamma_1(t) \rangle = 0 \\ \langle \Gamma(t)\Gamma(t') \rangle &= \delta(t-t') \\ \langle \Gamma_1(t)\Gamma_1(t') \rangle &= 2D\delta(t-t') \end{aligned} \tag{2.41}$$

In our analysis, we consider absence of correlation between Γ and Γ_1, hence

$$\langle \Gamma(t)\Gamma_1(t') \rangle = \langle \Gamma_1(t)\Gamma(t') \rangle = 0, \tag{2.42}$$

As discussed in [57], adiabatic elimination of ϵ is exact in the limit of $\tau \to 0$ and $\tau \to \infty$, as well as provides an adequate approximation for intermediate values of τ, which will eventually turn out pertinent to our interests, as described in next chapters. To begin with, we rearrange eq. 2.39 as,

$$\epsilon(x, \dot{x}) = \frac{1}{g_1(x)}(\dot{x} - h(x) - g_2(x)\Gamma(t)), \tag{2.43}$$

Evoking a mean field approximation, we can write,

$$\epsilon_{mf}(x, \dot{x}) = \frac{1}{g_1(x)}(\dot{x} - h(x)), \tag{2.44}$$

Next, we rescale time as $\hat{t} = \tau^{-1/2}t$ and take time derivative of ϵ_{mf},

$$\tau^{-1/2}\dot{\epsilon}_{mf} = \frac{\tau^{-1/2}}{g_1(x)}\left(h(x)\frac{g_1'(x)}{g_1(x)} - h'(x)\right)\dot{x} + \frac{\tau^{-1}}{g_1(x)}\left(\ddot{x} - \frac{g_1'(x)}{g_1(x)}\dot{x}^2\right), \tag{2.45}$$

In eq. 2.45, prime denotes derivation with respect to x and overdots are derivation with respect to $\hat{t}$. Imposing time scale separation gives us freedom to compare coefficients of $\tau^{-1/2}$ and τ^{-1} on both sides of eq. 2.45. Hence, we can write,

$$\ddot{x} - \frac{g_1'(x)}{g_1(x)}\dot{x}^2 = 0, \tag{2.46}$$

and,

$$\dot{\epsilon}_{mf} = \frac{\dot{x}}{g_1(x)}\left(h(x)\frac{g_1'(x)}{g_1(x)} - h'(x)\right), \tag{2.47}$$

Eventually, in eq. 2.39, we can replace ϵ with ϵ_{mf} using eq. 2.44. Similarly, in eq. 2.40, we can replace $\dot{\epsilon}$ with $\dot{\epsilon}_{mf}$ with the help of eq. 2.47. These transformations lead to a unified equation, combining eq. 2.39 and 2.40, also known as Unified Colored Noise Approximation.

$$\dot{x} = \frac{h(x)}{C(\tau,x)} + \frac{1}{C(\tau,x)}[g_1(x)\Gamma_1(t) + g_2(x)\Gamma(t)], \tag{2.48}$$

where $C(\tau,x) = 1 - \tau\left[h'(x) - \frac{g_1'(x)}{g_1(x)}h(x)\right]$. We can further transform eq. 2.48 as,

$$\dot{x} = \frac{h(x)}{C(\tau,x)} + \frac{1}{C(\tau,x)}g(x)\tilde{\Gamma}(t), \tag{2.49}$$

where,

$$g(x)\tilde{\Gamma}(t) = g_1(x)\Gamma_1(t) + g_2(x)\Gamma(t), \tag{2.50}$$

where $\tilde{\Gamma}$ is also δ-correlated Gaussian white noise similar to Γ and Γ_1.

$$\begin{aligned} \langle \tilde{\Gamma}(t) \rangle &= 0 \\ \langle \tilde{\Gamma}(t)\tilde{\Gamma}(t') \rangle &= 2\delta(t-t'), \end{aligned} \tag{2.51}$$

Square of both sides and taking expectation, eq. 2.50 transforms into,

$$\begin{aligned} g^2(x)\langle \tilde{\Gamma}(t)\tilde{\Gamma}(t') \rangle &= {g_1}^2(x)\langle \Gamma(t)\Gamma(t') \rangle + {g_2}^2(x)\langle \Gamma_1(t)\Gamma_1(t') \rangle \\ &+ g_1(x)g_2(x)\langle \Gamma(t)\Gamma(t') \rangle + g_2(x)g_1(x)\langle \Gamma(t')\Gamma(t) \rangle, \end{aligned} \tag{2.52}$$

As mentioned in eq. 2.42, we consider absence of correlation between Γ and Γ_1, thus using correlators of Γ, Γ_1 from eq. 2.41 and $\tilde{\Gamma}_1$ from eq. 2.51, we can write,

$$g(x) = \sqrt{D{g_1}^2(x) + \frac{1}{2}{g_2}^2(x)}, \tag{2.53}$$

Thus, an equivalent transformed SDE can be written as,

$$\dot{x} = \frac{h(x)}{C(\tau, x)} + \frac{1}{C(\tau, x)} g(x)\tilde{\Gamma}(t), \tag{2.54}$$

Due to initial finite correlation time scales, Eq. 2.47 is a Stratonovich SDE, where corresponding FPE includes a correction in drift term. We can write the FPE as,

$$\frac{dP(x,t)}{dt} = -\frac{\partial}{\partial x}A(x)P(x,t) + \frac{\partial^2}{\partial x^2}B(x)P(x,t) \tag{2.55}$$

where,

$$A = \frac{h(x)}{C(x)} + \frac{g'(x)}{C(x)}\left[\frac{d}{dx}\left(\frac{g(x)}{C(x)}\right)\right] \text{ and } B = \left(\frac{g(x)}{C(x)}\right)^2,$$

For Eq. 2.55, we can write a stationary state probability distribution as,

$$P_S(x) = N\frac{1}{B(x)}exp\left(\int \frac{A(x)}{B(x)}dx\right) \tag{2.56}$$

Similarly, distribution modes can be computed by putting first derivative of $P_S(x)$ equals to zero which leads to,

$$A(x) = B'(x) \tag{2.57}$$

Once again, it is important to point out here that, unlike SSA, CFPE/CLE or LNA, applications of UCNA are context specific and so far there have been only a few instances where UCNA or its variations have been applied as analytical tools in computational or systems biology [73,74].

Chapter 3

EFFECTS OF EXTERNAL FLUCTUATIONS ON STOCHASTIC BIMODALITY

3.1 Introduction

In this chapter, we will examine effects of external fluctuations on reaction networks, in the form of transitions in stationary state stochastic bimodality. Bimodality in biology appears as either existence of population wide multiple phenotypes or temporal phenotypic switching [14, 59, 69, 73] also known as regime shifts. It occurs due to cellular decision making eventually influenced by the topology of gene regulatory networks. It is typically marked by bimodal distributions of monitored characteristics such as cell size distributions, pathogenic virulence potency or protein expression levels in isogenic cellular populations [75–79]. Each mode is associated with a particular phenotype. We define external fluctuation as temporal discrete signal, which influences reaction propensities and demonstrates shift in system's original phenotype, i.e. most of the mass of bimodal distribution switches from being largely around one mode to another. We also discuss simplistic analytical approximations to transform our two-variable model to a one-variable model and applications of UCNA to capture these transitions.

Our results cover comparison between SSA simulations and UCNA based analytical approximation, along with sensitivity analysis through analytical formulation.

3.2 Schlögl Model

First proposed in 1972 [80], Schlögl model represents a single species, auto-catalytic chemical reaction system,

$$\begin{aligned} A + 2X &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} 3X \\ B &\underset{k_4}{\overset{k_3}{\rightleftharpoons}} X \end{aligned} \tag{3.1}$$

here, A and B represent reservoir species with constant concentrations a and b, also known as pump parameters. Hence, Schlögl model is an open system. Following simple mass-balance, we can write RE for deterministic kinetics of x, concentration of X as,

$$\frac{dx}{dt} = k_1 a x^2 - k_2 x^3 + k_3 b - k_4 x, \tag{3.2}$$

As a and b are constant model parameters, for the sake of conciseness, we can re-write Eq. 3.2 with a new time scale $t' = k_4 t$ as,

$$\frac{dx}{dt'} = k_1' x^2 - k_2' x^3 + k_3' - x, \tag{3.3}$$

where, $k_1' = \frac{k_1 a}{k_4}$, $k_2' = \frac{k_2}{k_4}$ and $k_3' = \frac{k_3 b}{k_4}$. Following Eq. 3.3, Schlögl model exhibits a bifurcation based on discriminant Δ of Eq. 3.3,

$$\Delta = {k_1'}^2 - 4k_2' - 4{k_1'}^3 k_3' + 18 k_1' k_2' k_3' - 27 {k_2'}^2 {k_3'}^2, \tag{3.4}$$

For $\Delta > 0$, reaction system will have three distinct roots, thus it will exhibit bistability. For $\Delta = 0$, there will be only one repeated real root. For $\Delta < 0$, there will be one real root along with two non real complex conjugate roots, hence system stays in monostable regime. Fig. 3.1 plots implicit function $\Delta = 0$ with respect to k_1' and k_3' (hence, with respect to inbuilt parameter a and b) and shows enclosed bistable regime.

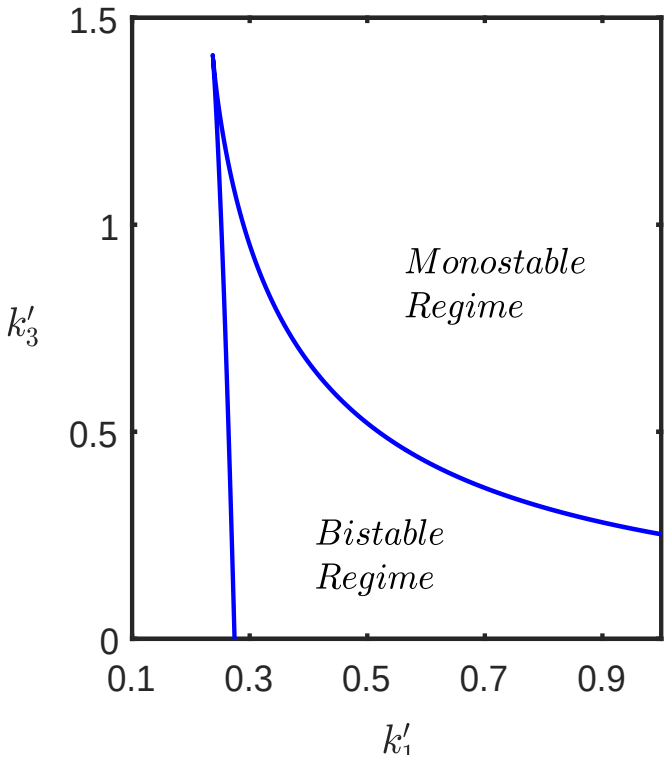

Figure 3.1: Mono and bistable regimes of Schlögl model corresponding to one and three real roots respectively, in k_1'-k_3' plane for $k_2' = 0.0188$. Blue line is a plot of discriminant $\Delta = 0$, as shown in Eq. 3.4.

Fig. 3.2 provides a further detailed picture of Schlögl model's bistable kinetics. For different values of parameter k_1' and initial conditions, x attains different stationary state.

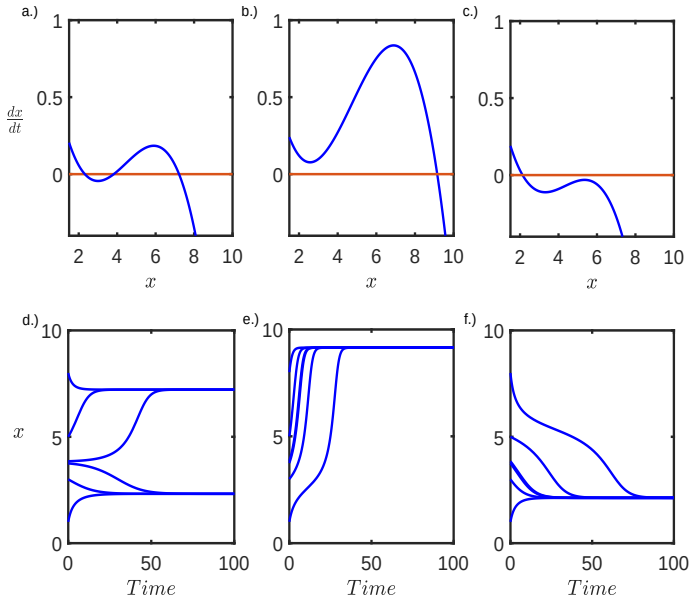

Figure 3.2: Plots of RE from Eq. 3.3 and simulated deterministic trajectories with different initial conditions for three different sets of parameters. (a), (b) and (c) show number of roots of RE. (d), (e) and (f) show simulated deterministic trajectories of x. For (a) and (d) $k_1' = 0.251$, for (b) and (e) $k_1' = 0.267$ and for (c) and (f) $k_1' = 0.244$. $k_2' = 0.0188$ and $k_3' = 1.203$

According to deterministic kinetics, we see macroscopic concentrations following trajectories depend upon initial concentration and macroscopic reaction rate constants. However, mesoscopic kinetics, considering system state as a finite number of reacting molecules due to a finite system size, leads to a different story [58, 60, 81]. To start with, we can write CME for reaction system in Eq. 3.1 as,

$$\frac{dP(n,t)}{dt} = \left(k_1 a \frac{(n-1)(n-2)}{\Omega} + k_3 b \Omega\right) P(n-1,t) + \left(k_2 \frac{(n+1)n(n-1)}{\Omega^2} .. \right.$$
$$\left. .. + k_4(n+1)\right) P(n+1,t) - \left(k_1 a \frac{n(n-1)}{\Omega} + k_3 b \Omega .. \right.$$
$$\left. .. + k_2 \frac{n(n-1)(n-2)}{\Omega^2} + k_4 n\right) P(n,t), \quad (3.5)$$

here, n is system state referring to molecules of X and Ω is volume of a well-stirred reaction system. While CME in Eq. 3.5 is highly non-linear in nature, an approximated stationary state solution is provided using detailed balance. To start with, following notation from [8], we can write a stationary state

solution for Eq. 3.5 as,

$$0 = J(n+1) - J(n), \tag{3.6}$$

where,

$$J(n) = w^-(n)P_s(n) - w^+(n-1)P_s(n-1), \tag{3.7}$$

$$w^+(n) = \left(k_1 a \frac{n(n-1)}{\Omega} + k_3 b\Omega\right), \tag{3.8}$$

$$w^-(n) = \left(k_2 \frac{n(n-1)(n-2)}{\Omega^2} + k_4 n\right), \tag{3.9}$$

here, $J(n)$ is defined as probability current of discrete jump process. $P_s(n)$ is stationary state probability of system being in state n and $w^+(n), w^-(n)$ are transition probability from state n to $(n+1)$ and to state $(n-1)$ respectively. As we are dealing with molecule numbers, which are always positive integers, hence it is obvious to write,

$$\begin{aligned} w^-(n) &= 0, \quad \text{for} \quad n \leq 0, \\ w^+(n) &= 0, \quad \text{for} \quad n < 0, \\ P_s(n) &= 0, \quad \text{for} \quad n < 0, \end{aligned}$$

thus, we can write,

$$J(0) = w^-(0)P_s(0) - w^+(-1)P_s(-1) = 0, \tag{3.10}$$

Now, summing up Eq. 3.6, using Eq. 3.10, we can write,

$$0 = \sum_{j=1}^{n-1} [J(j+1) - J(j)] = J(n) - J(0) = J(n),$$

hence,

$$J(n) = 0, \tag{3.11}$$

Using Eq. 3.11 and 3.7, we can write,

$$P_s(n) = \frac{w^+(n-1)}{w^-(n)} P_s(n-1), \tag{3.12}$$

or,

$$P_s(n) = P_s(0) \prod_{j=1}^{n} \frac{w^+(j-1)}{w^-(j)}, \tag{3.13}$$

While Eq. 3.13 provides an exact recursive solution for Schlögl model following detailed balance properties, yet it remains computationally expensive. Hence, another approximation, which further simplifies this solution, remains applicable during large system size conditions. Following Eq. 3.13, it can be transformed as,

$$P_s(n) = P_s(0) exp\bigg(- \sum_{j=1}^{n} ln \frac{w^-(j)}{w^+(j-1)} \bigg), \tag{3.14}$$

We can simplify Eq. 3.14 by putting Euler-McLauring summation formula,

$$\sum_{j=1}^{n} ln \frac{w^-(j)}{w^+(j-1)} \approx \int_1^n ln \frac{w^-(j)}{w^+(j-1)} dj + \frac{1}{2} \bigg(ln \frac{w^-(n)}{w^+(n-1)} + ln \frac{w^-(1)}{w^+(0)} \bigg). \tag{3.15}$$

As shown in Eq. 3.15, Euler-McLaurin summation formula approximates

discrete system state to a continuous variable. Thus, for the sake of consistency, we can transform system state probability distribution $P_s(n)$ to concentration probability distribution $P_s(x)$ using $x = \frac{n}{\Omega}$ and change of variable $P_s(x) = \Omega P_s(n)$. Hence, for sufficiently large system size $(n \gg 1)$, Eq. 3.14 can be written as,

$$P_s(x) = N(x)exp\bigg(-\Omega \int_0^x ln\frac{r^-(x)}{r^+(x)}dx\bigg), \tag{3.16}$$

where,

$$r^+(x) = (k_1 a x^2 + k_3 b), \tag{3.17}$$

$$r^-(x) = (k_2 x^3 + k_4 x), \tag{3.18}$$

In Eq. 3.16, lower integration limit is changed to zeros, assuming $\frac{1}{\Omega} \approx 0$, for large system size. Solving Eq. 3.16 provides an analytical formulation for $P_s(x)$,

$$P_s(x) = N(x)exp\big(-\Omega\Phi(x)\big), \tag{3.19}$$

where, $\Phi(x)$ denotes stochastic potential in large system size limit,

$$\begin{aligned}\Phi(x) \;\; &= -2\sqrt{\tfrac{bk_3}{ak_1}}ArcTan\bigg(\sqrt{\tfrac{ak_1}{bk_3}}x\bigg) + 2\sqrt{\tfrac{k_4}{k_2}}ArcTan\bigg(\sqrt{\tfrac{k_2}{k_4}}x\bigg) \\ &\quad -x\bigg(1 + ln\bigg(\tfrac{bk_3 + ak_1x^2}{k_4x + k_2x^3}\bigg)\bigg),\end{aligned} \tag{3.20}$$

and prefactor $N(x)$ with normalization constant Z,

$$N(x) = \frac{k_2x^2 + k_4}{Z\sqrt{x}\bigg(x^2 + \frac{k_4 b}{k_1 a}\bigg)} \tag{3.21}$$

As shown in Fig. 3.3, recursive solution (Eq. 3.13) based on detailed balance properties of Schlögl model matches accurately with SSA results

independent of system sizes. Contrarily, approximate solution (Eq. 3.19) seems to overestimate the distribution mode corresponding to higher concentration. Its accuracy increases as system size increases. In Fig. 3.3 and in further analysis, parameters are chosen to keep analysis consistent with deterministic re-parametrized Eq. 3.3.

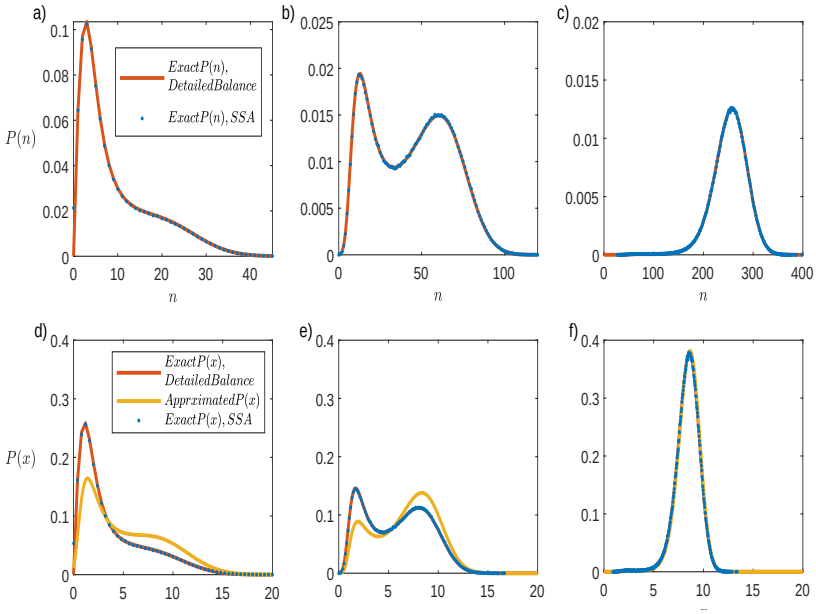

Figure 3.3: Change in stationary state distribution bimodality for Schlögl model, due to change in system size Ω. Solid red line corresponds to recursive detailed balance solution from Eq. 3.13. Solid yellow line denotes approximated solution from Eq. 3.19. Blue dots refer to probability distribution captured using SSA. $P(n)$ and $P(x)$ refer to probability distribution of system state as molecule numbers and as concentration, respectively. For (a) and (d) $\Omega = 2.5$, for (b) and (e) $\Omega = 7.5$ and for (c) and (f) $\Omega = 30$. Other model parameters are $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, a = 1$ and $b = 1$

As evident in Fig. 3.3, an interesting aspect of mesoscopic behaviour is the transition of probability distribution from being unimodal to bimodal and again to unimodal, with increasing system size [81]. This transition can be defined as stochastic bifurcation. Using detailed balance, we can show stochastic bifurcation dependence upon system size and model parameters. It is quite intuitive that at stationary state, transition probability of moving from state $(j-1)$ to j will be equal to transition probability of moving from state j to $(j-1)$. Exactly same condition can be derived using equivalence of deterministic and stochastic stationary states as presented in [8]. We can

write the deterministic rate equation for system state n as,

$$\frac{dn}{dt} = w^+(n) - w^-(j), \tag{3.22}$$

for which stationary state condition results in,

$$w^+(n) = w^-(j-1), \tag{3.23}$$

which lead to maximum in $P_s(n)$, when

$$P_s(n) \approx P_s(n-1), \tag{3.24}$$

which again following eq. 3.12 leads us to,

$$w^+(j-1) = w^-(j), \text{ for } j \in [1, n] \tag{3.25}$$

hence, using transformation of variables with Eq. 3.8 and 3.9, we can further expand the eq. 3.25 for $j = n$ as,

$$F\Omega\left(\frac{(n-1)(n-2)}{\Omega^2} + G\right) = n\left(\frac{(n-1)(n-2)}{\Omega^2} + H\right), \tag{3.26}$$

where, $F = \frac{k_1 a}{k_2}$, $G = \frac{k_4}{k_2}$ and $H = \frac{k_3 b}{k_1 a \Omega}$. Eq. 3.26 can be further simplified as,

$$-n^3 + (\nu + 3)n^2 - (3\nu + 2 + \beta\nu^2)n + (2\nu + \gamma\nu^3) = 0, \tag{3.27}$$

where, $\nu = F\Omega$, $\beta = \frac{G}{F^2}$ and $\gamma = \frac{H}{F^2}$. Eq. 3.27 is a cubic equation similar to deterministic RE from Eq. 3.2, hence the sign of discriminant Δ' reveals exact nature of distribution modality. Thus, plotting $\Delta' = 0$ from Eq. 3.28 depicts

parameter regimes in form of transformed variables ν, β and γ.

$$\Delta' = \left(\frac{\beta}{3} - \frac{1}{9} + \frac{1}{3\nu} - \frac{1}{3\nu^2}\right)^3 + \left(-\frac{1}{27} + \frac{\beta}{6} - \frac{\gamma}{2} + \frac{1}{2\nu}\left(\beta + \frac{1}{3}\right) - \frac{1}{6\nu}\right)^2 = 0, \tag{3.28}$$

From above equation, a solution for γ can also be written as,

$$\gamma(\beta, \nu) = \frac{1}{3}\left(\beta - \frac{2}{9}\right) + \frac{1}{\nu}\left(\beta + \frac{1}{3}\right) - \frac{1}{3\nu^2} \pm \frac{2}{3\sqrt{3}}\left(\frac{1}{3} - \beta - \frac{1}{\nu} + \frac{1}{\nu^2}\right)^{\frac{3}{2}}, \tag{3.29}$$

Hence, using explicit solution for γ as a function of β and ν, we can also write for cusp of the plot in Fig. 3.4 as,

$$\beta_c = \frac{1}{3} - \frac{1}{\nu} + \frac{1}{\nu^2}, \tag{3.30}$$

and

$$\gamma_c = \frac{1}{27} + \frac{1}{3\nu} - \frac{1}{\nu^2} + \frac{1}{\nu^3}, \tag{3.31}$$

In Eq. 3.29, ν is the only parameter dependent upon system size, as given previously $\nu = F\Omega$, hence for $\Omega \to \infty$, we can write,

$$\gamma_\infty = \frac{1}{3}\left(\beta - \frac{2}{9}\right) \pm \frac{2}{3\sqrt{3}}\left(\frac{1}{3} - \beta\right)^{\frac{3}{2}}, \tag{3.32}$$

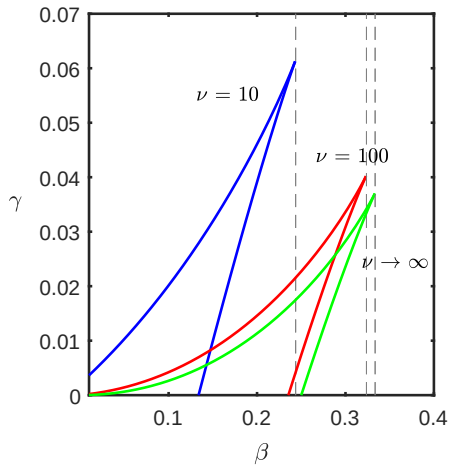

Figure 3.4: Stochastic phase plane of Schlögl model with respect to transformed parameters β, γ and ν, using Eq. 3.31 which represents solution of discriminant $\Delta' = 0$ (Eq. 3.30). Enclosed space represents bimodality regime with three real roots of Eq. 3.27 for three different system sizes. Here $\nu = F\Omega$ indicates change in system size. $\beta = \frac{G}{F^2}$ and $\gamma = \frac{H}{F^2}$ with $F = \frac{k_1 a}{k_2}$, $G = \frac{k_4}{k_2}$ and $H = \frac{k_3 b}{k_1 a \Omega}$.

As shown in Fig. 3.4, parametric space for bimodality demonstrates a shift according to transformed variable ν which is directly proportional to system size Ω. Fig. 3.4 is another example of stark differences from macroscopic dynamics due to inherent molecular fluctuations of reacting species. It is clear that with change in system size Ω, reaction system is capable of exhibiting both unimodal and bimodal dynamics for same set of macroscopic rate parameters.

3.2.1 CFPE Application to Schlögl Model

As mentioned chapter 2, CFPE is one of the most popular approximations of CME, specially for one-variable models. For analytical treatment of Schlögl model, CFPE has been implemented successfully and provides analytical details of mesoscopic bimodality. While it has been shown that stochastic potential derived using CFPE is different than the one derived using detailed balance and Euler-McLauring approximation as in Eq. 3.20, even when $\Omega \to \infty$ [58], yet CFPE along with CLE still finds broad applicability because

of simplicity of analysis [69].

To begin our analysis, CFPE corresponding to reaction scheme given in Eq. 3.1 can be written as,

$$\frac{dP(n,t)}{dt} = -\frac{\partial}{\partial n}A(n)P(n,t) + \frac{1}{2}\frac{\partial^2}{\partial n^2}B(n)P(n,t), \tag{3.33}$$

where,

$$A(n) = k_1 a\frac{n^2}{\Omega} - k_2\frac{n^3}{\Omega^2} + k_3 b\Omega - k_4 n, \tag{3.34}$$

constitutes *drift* term of CFPE, similarly

$$B(n) = k_1 a\frac{n^2}{\Omega} + k_2\frac{n^3}{\Omega^2} + k_3 b\Omega + k_4 n, \tag{3.35}$$

constitutes *diffusion* term of CFPE. We can write stationary state probability distribution of system state n as,

$$P_s(n) = \frac{\mathcal{N}}{B(n)}exp\left[\int_0^n \frac{2A(n')}{B(n')}dn'\right], \tag{3.36}$$

where $\mathcal{N}$ is normalization constant. Molecule number n represents system state as discrete positive integers, hence for the sake of consistency, it is appropriate to transform the probability distribution in Eq. 3.36, in the form of concentration distribution. Hence, using $x = \frac{n}{\Omega}$ and $P(x) = \Omega P(n)$. Eq. 3.36 can be transformed as,

$$P_s(x) = \frac{\mathcal{N}'}{B(x)}exp\left[\Omega\int_0^x \frac{2A(x')}{B(x')}dx'\right], \tag{3.37}$$

where $\mathcal{N}'$ is normalization constant for transformed distribution $P_s(x)$ and

assuming sufficiently large system size, we can write,

$$A(x) = k_1 a x^2 - k_2 x^3 + k_3 b - k_4 x, \tag{3.38}$$

and

$$B(x) = k_1 a x^2 + k_2 x^3 + k_3 b + k_4 x, \tag{3.39}$$

Similarly, extremum of distribution in Eq. 3.37 can be derived by equating the first derivative to zero, which leads to,

$$A(x) = \frac{1}{2\Omega} B'(x), \tag{3.40}$$

In Fig. 3.5, accuracy of CFPE for sufficiently large system size is evident. Concentration distribution captured using SSA simulations for different values of k_1 agrees well with approximated concentration distribution captured using CFPE, hence proving utility of CFPE even for highly non-linear reaction network. Fig. 3.5 shows a transition of distribution weight around different modes with change in parameters. In biological context, it can be understood as distribution of multiple phenotypes. While in Fig. 3.5, unimodal and bimodal distributions are associated with unique and multiple real valued deterministic stationary states, respectively, yet as shown previously (Fig. 3.4) that it is not always true. For certain parameter values, bimodal mesoscopic distributions can prevail while there is only single deterministic steady state and *vice versa*.

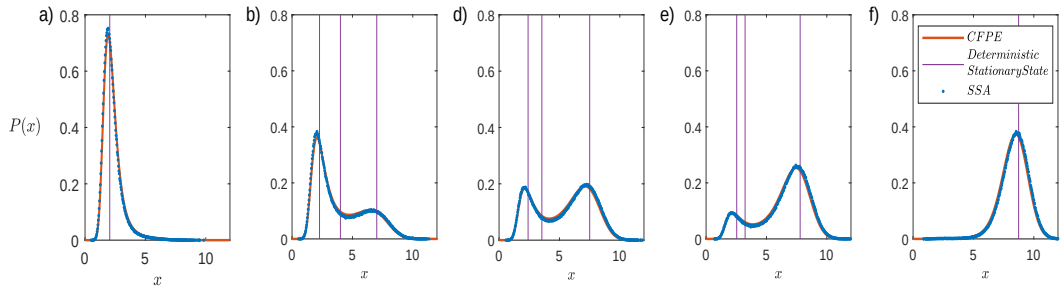

Figure 3.5: For sufficiently large system sizes, analytical distribution $P(x)$ derived using CFPE shows remarkable agreement with the one extracted from Schlögl model SSA simulations, for different values of k_1 (a) 0.240, (b) 0.250, (c) 0.253, (d) 0.255, (e) 0.263. Other parameters values fixed to: $k_2 = 0.0188, k_3 = 1.203, k_4 = 1, a = 1, b = 1, \Omega = 30$. Vertical bars indicate deterministic stationary state for each set of parameters.

3.3 Effects of External Fluctuations on Mesoscopic Bimodality

In a cellular context, reaction networks associated with gene expression and regulations function in a dynamic environment, which finds parallel with our concept of missing intermediates as source of external fluctuations: Reaction networks only portray simplistic mechanisms of central dogma and skip most of the intermediary species as well as catalysing enzyme molecules. Thus, there is always scope for including temporal environmental fluctuations which influence reaction propensities as well as mimic dynamics of intermediate transient molecules.

To begin with, we study effects of external fluctuations on mesoscopic bimodality of Schlögl model. We included an extrinsic species Y with macroscopic concentration y and molecule number n_y, following an independent birth-death mechanism,

$$\Phi \underset{b_1}{\overset{b_o}{\rightleftharpoons}} Y \tag{3.41}$$

Hence, RE for Y can be written as,

$$\frac{dy}{dt} = b_o - b_1 y, \tag{3.42}$$

As generation and decay of Y is independent of kinetics of X altogether, hence it is convenient to write both macroscopic and mesoscopic kinetics for Y beforehand. Solving rate Eq. 3.42 with initial condition $y(t = 0) = y_o$, gives us,

$$y = \frac{bo}{b1}(1 - exp^{(-b_1 t)}) + y_o exp^{(-b_1 t)}, \tag{3.43}$$

For $t \to \infty$, we get stationary state concentration y_s as,

$$y_s = \frac{bo}{b1}, \tag{3.44}$$

Similarly, we can write CME for reaction scheme shown in Eq. 3.41 as,

$$\frac{dP(n_y, t)}{dt} = \Big(b_o \Omega P((n_y - 1), t) + b_1 (n_y + 1) P((n_y + 1), t) \Big) ... \\ ... - \Big(b_o \Omega P(n_y, t) + b_1 n_y P(n_y, t) \Big), \tag{3.45}$$

As provided in [8], it is possible to derive an exact time dependent solution of $P(n_y, t)$, but for the sake of simplicity of analysis and relevance with following examples, only stationary state solution is provided. To begin with, we can transform CME in Eq. 3.45, using probability generating function,

$$G(s, t) = \sum_{n_y=0}^{\infty} s^{n_y} P(n_y, t), \tag{3.46}$$

hence, transformed CME can be written as,

$$\frac{\partial G(s,t)}{\partial t} = b_o\Omega(s-1)G(s,t) - b_1(s-1)\frac{\partial G(s,t)}{\partial s}, \tag{3.47}$$

For stationary solution of Eq. 3.47, we put right hand side to zero and replace $G(s,t)$ with $G(s)$,

$$b_o\Omega G(s) - b_1\frac{\partial G(s)}{\partial s} = 0, \tag{3.48}$$

Using Eq. 3.46, we can write,

$$G(s=1) = \sum_{n_y=0}^{\infty} P(n_y,t) = 1, \tag{3.49}$$

Hence, integrating Eq. 3.48 from $s' = 1$ to $s' = s$ for $G(s'=1) = 1$ to $G(s)$, yields,

$$G(s) = exp^{\frac{b_o\Omega}{b_1}(s-1)} = exp^{-\frac{b_o\Omega}{b_1}} exp^{\frac{b_o\Omega}{b_1}s}, \tag{3.50}$$

which using series expansion, can be written as,

$$G(s) = exp^{-\frac{b_o\Omega}{b_1}} \sum_{n_y=0}^{\infty} \left(\frac{b_o\Omega}{b_1}\right)^{n_y} \frac{s^{n_y}}{n_y!}, \tag{3.51}$$

Comparing Eq. 3.51 with Eq. 3.46 renders stationary state distribution $P_s(n_y)$ as,

$$P_s(n_y) = exp^{-\frac{b_o\Omega}{b_1}} \left(\frac{b_o\Omega}{b_1}\right)^{n_y} \frac{1}{n_y!} \tag{3.52}$$

Same solution can also be obtained using detailed balance properties. Distribution in Eq. 3.52 denotes a Poisson distribution, with equal mean and variance along with exponential autocorrelation function. In this context, it is

convenient to mention moment equations, hence using CME given in Eq. 3.45, we can write, equation for first moment (mean),

$$\frac{d\langle n_y(t)\rangle}{dt} = b_o\Omega - b_1\langle n_y\rangle, \tag{3.53}$$

which turns out to be the same as deterministic rate Eq. 3.42. Similarly, equation for second centralized moment (variance), using Eq. 3.45 and 3.53, is,

$$\frac{d\langle n_y(t)^2\rangle}{dt} - \frac{d\langle n_y(t)\rangle^2}{dt} = \frac{d\langle n_y(t)^2\rangle}{dt} - 2\langle n_y(t)\rangle\frac{d\langle n_y(t)\rangle}{dt} = b_o\Omega + b_1\langle n_y\rangle, \tag{3.54}$$

Hence, we can write,

$$\langle n_y\rangle_s = \frac{b_o\Omega}{b_1}, \tag{3.55}$$

$$var[n_y]_s = \frac{b_o\Omega}{b_1}, \tag{3.56}$$

$$\langle n_y(t)n_y(0)\rangle_s = \frac{b_o\Omega}{b_1}exp^{-b_1t}, \tag{3.57}$$

While mean and variances are evident from Eq. 3.52 as well as from 3.53 and 3.54, to compute autocorrelation function, Regression Theorem for linear systems is utilised [8]. Hence, with a Poisson distribution and finite time, exponential autocorrelation function, Y represents a source of Coloured Poisson Noise [82].

While birth-death mechanism of external species Y is exactly solvable even in mesoscopic context, it is important to present Langevin equation associated with LNA and its differences from exact solution. This approximate analysis

will set the platform for implementing UCNA analysis under environmental fluctuations.

As shown in chapter 2, LNA fits a Normal distribution for inherent molecular fluctuations ε_{n_y} around deterministic mean. Hence, we can start with FPE derived using LNA for reaction system in Eq. 3.41 as,

$$\frac{\partial \Pi(\varepsilon_{n_y}, t)}{\partial t} = b_1 \frac{\partial}{\partial \varepsilon_{n_y}}(\varepsilon_{n_y} \Pi(\varepsilon_{n_y}, t)) + b_o \frac{\partial^2}{\partial {\varepsilon_{n_y}}^2} \Pi(\varepsilon_{n_y}, t) \tag{3.58}$$

We can write corresponding SDE as,

$$d\varepsilon_{n_y} = -b_1 \varepsilon_{n_y} dt + \sqrt{2b_o} dW_t, \tag{3.59}$$

for which corresponding stationary state distribution can be written as,

$$\Pi_s(\varepsilon_{n_y}) = \frac{1}{\sqrt{2\pi(\frac{b_o}{b_1})}} exp^{-\frac{1}{2}\frac{b_1}{b_o}{\varepsilon_{n_y}}^2}, \tag{3.60}$$

Thus, using ansatz $n_y = y\Omega + \varepsilon_{n_y}\sqrt{\Omega}$, an SDE and stationary state distribution for molecule number n_y, approximating it as a continuous variable, can be written as,

$$dn_y = -b_1(n_y - \frac{b_o \Omega}{b_1})dt + \sqrt{2b_o\Omega} dW_t, \tag{3.61}$$

for which corresponding stationary state distribution is,

$$P_s(n_y) = \frac{1}{\sqrt{2\pi(\frac{b_o\Omega}{b_1})}} exp^{-\frac{1}{2}\frac{b_1}{b_o\Omega}(n_y - \frac{b_o\Omega}{b_1})^2}, \tag{3.62}$$

SDE in Eq. 3.61 corresponds to the OU process, which has the same mean, variance and autocorrelation function but different distribution, in comparison to the exact solution derived using generating function. Fig. 3.6,

compares distribution in Eq. 3.52 with 3.62. It is evident that with increasing mean, Poisson distribution and discrete Markovian jump process can be approximated with Normal distribution and continuous Markovian diffusion with sufficient accuracy.

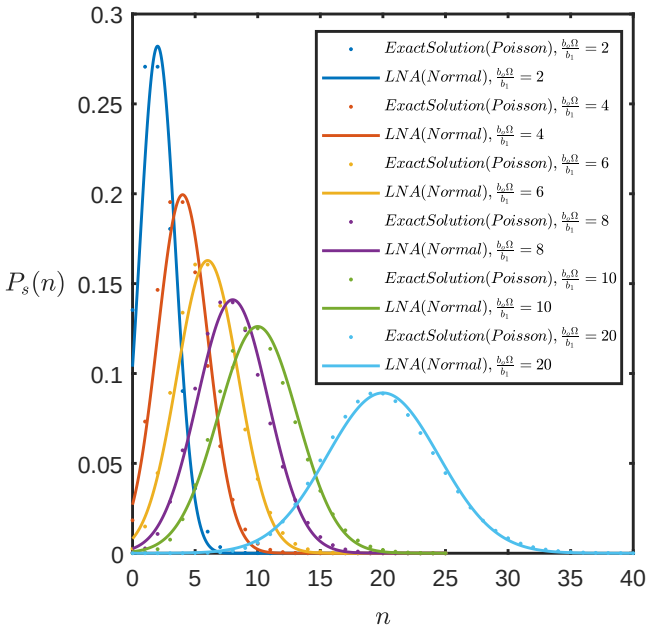

Figure 3.6: Agreement between LNA derived stationary state distribution (Normal) and exact distribution derived using CME (Poisson) increases with the increase in $\frac{b_0\Omega}{b_1}$, the stationary state mean number of molecules of Y which follows simple birth-death mechanism with birth rate b_o, death rate b_1 and system size Ω.

Now, simplest form of interactions of species Y with dynamics of Schlögl model can be defined as a catalytic molecule, where reaction rates depend upon concentration y but Y don't change reaction stoichiometry itself. Hence, based on reaction influenced by external fluctuations, we can write multiple altered reaction schemes as following,

Reaction Set	
R_1	$A + 2X + Y \xrightarrow{\frac{1}{y_s}k_1} 3X + Y$
R_2	$3X + Y \xrightarrow{\frac{1}{y_s}k_2} A + 2X + Y$
R_3	$B + Y \xrightarrow{\frac{1}{y_s}k_3} X + Y$
R_4	$X + Y \xrightarrow{\frac{1}{y_s}k_4} B + Y$

For each reaction set, individual reactions are modified and the rest of the reactions remain intact, same as in original reaction scheme in Eq. 3.1. For each modified reaction, macroscopic reaction rate is divided by stationary state concentration of Y. This re-parametrization, makes sure that deterministic stationary states of each reaction set, remains exactly the same as the deterministic stationary states of original Schlögl model from Eq. 3.1, though inclusion of Y will affect time scales to reach stationary state. For example, for reaction set R_1, we can write RE as,

$$\frac{dx}{dt} = \frac{1}{y_s}k_1 a x^2 y - k_2 x^3 + k_3 b - k_4 x, \tag{3.63}$$

From mesoscopic perspective, we are interested in effects of inclusion of this dynamic species on distribution bimodality. To start our investigation, we performed SSA simulations and produced stationary state distribution for above mentioned reaction sets. Fig. 3.7 shows transition in distribution weights from higher to lower mode with decreasing birth-rate of Y (b_o), corresponding to increasing external fluctuation intensity, for fixed death-rate (b_1). It is apparent that discrete environmental fluctuations can force stationary distribution toward lower mode, or in biological context toward a *basal* state. Sensitivity of these transitions is negligible for reaction

set R_3, where fluctuations of Y influences propensity of zero order birth reaction and different for different reaction sets, as shown later on.

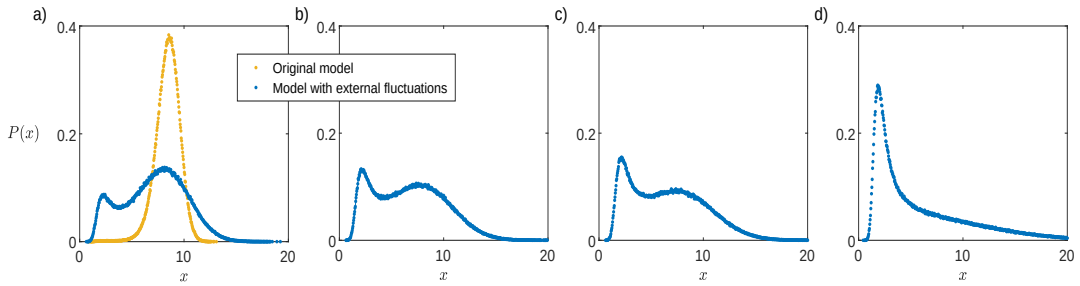

Figure 3.7: Transition in stationary state bimodal distribution of X concentration, following reaction set R_1. Yellow dots are stationary state distribution of X in original Schlögl model and blue dots are stationary state distribution of X due to interactions with Y in reaction set R_1, both extracted from SSA simulation trajectories. Distribution mass shifts from higher to lower mode with increase in the intensity of external fluctuations from Y, i.e. decreasing b_o. Distributions are extracted from single SSA trajectory. $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, \Omega = 30, b_1 = 1$. $b_o\Omega$ is (a) 80, (b) 50, (c) 40 and (d) 10.

For analytical prediction of these mesoscopic transitions, we reduce two-variable discrete jump process model to a one-variable continuous diffusion process. For example, for reaction set R_1,

$$\begin{aligned} A + 2X + Y &\xrightarrow{\frac{1}{y_s}k_1} 3X + Y \\ 3X &\xrightarrow{k_2} 2X + A \\ B &\underset{k_4}{\overset{k_3}{\rightleftharpoons}} X \\ \Phi &\underset{b_1}{\overset{b_o}{\rightleftharpoons}} Y \end{aligned} \tag{3.64}$$

we can write propensity for modified reaction as,

$$w = \frac{1}{y_s}k_1\frac{n(n-1)n_y}{\Omega^2}, \tag{3.65}$$

where, system state $\boldsymbol{n}$ consists of n molecules of X and n_y molecules of Y. Using previously shown approximation of discrete jump dynamics of Y as a

continuous diffusion process, we can approximate reaction set R_1 as,

$$\begin{aligned} A + 2X &\underset{k_2}{\overset{k_1(1+\zeta(t))}{\rightleftharpoons}} 3X \\ B &\underset{k_4}{\overset{k_3}{\rightleftharpoons}} X \end{aligned} \tag{3.66}$$

and write approximated propensity function as,

$$w' = k_1 \frac{n(n-1)}{\Omega}(1 + \xi(t)), \tag{3.67}$$

where, $\xi(t)$ is a *coloured* noise following,

$$(1 + \xi(t)) = \frac{1}{y_s}\frac{n_y}{\Omega}, \tag{3.68}$$

hence,

$$d\xi = \frac{1}{y_s}\frac{dn_y}{\Omega}, \tag{3.69}$$

which using Eq. 3.61 and $y_s = \frac{b_o}{b_1}$, can be written as,

$$d\xi = -b_1 \xi dt + \sqrt{\frac{2}{b_o \Omega}} b_1 dW_t, \tag{3.70}$$

Similar to Eq. 3.61, Eq. 3.70 represents another standard OU process, with stationary state mean, variance and autocorrelation function as,

$$\langle \xi(t) \rangle_s = 0, \tag{3.71}$$

$$var[\xi(t)]_s = \frac{b_1}{b_o \Omega}, \tag{3.72}$$

$$\langle \xi(t)\xi(0)\rangle_s = \frac{b_1}{b_o\Omega} exp^{-\frac{t}{\tau}}, \tag{3.73}$$

where, $\tau = \frac{1}{b_1}$ denotes time scale of random variable $\xi(t)$. Hence, for reaction scheme in Eq. 3.73, using CFPE for sufficiently large molecule number n, SDE based on a simplistic analytical formulation can be written as,

$$\begin{aligned} dn = \left(k_1 a\frac{n^2}{\Omega}(1+\xi) - k_2\frac{n^3}{\Omega} + k_3 b\Omega - k_4 n\right)dt... \\ ... + \sqrt{\left(k_1 a\frac{n^2}{\Omega}(1+\xi) + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}dW_t', \end{aligned} \tag{3.74}$$

An ad hoc transformation of SDE 3.74 for small noise limits ($var[\xi(t)]_s$) can be done as following,

Let's start with diffusion term of Eq. 3.74,

$$f(\delta) = \sqrt{\left(k_1 a\frac{n^2}{\Omega}(1+\delta\xi') + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}dW_t', \tag{3.75}$$

where,

$$\xi(t) = \delta\xi'(t) \tag{3.76}$$

hence,

$$var[\xi(t)]_s = \delta^2 = \frac{b_1}{b_o\Omega}, \tag{3.77}$$

Using Taylor's series expansion,

$$f(\delta) \approx f(0) + f'(0)\delta, \tag{3.78}$$

here,

$$f(0) = \sqrt{\left(k_1 a\frac{n^2}{\Omega} + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}dW_t', \tag{3.79}$$

and

$$f'(0) = \frac{1}{2\sqrt{\left(k_1 a\frac{n^2}{\Omega} + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}}k_1 a\frac{n^2}{\Omega}\xi' dW_t', \tag{3.80}$$

Using exact solution of OU process from Eq. 3.70, for transformed variable $\xi'(t)$, we can write,

$$\xi'(t) = \xi_o' exp^{-b_1 t} + \frac{1}{\delta}\sqrt{\frac{2}{b_o\Omega}}\int_0^t exp^{-b_1(t-s)}dW_s, \tag{3.81}$$

For $t \to \infty$, contribution of initial condition ξ_o' will eventually vanish, hence we can write,

$$\xi'_{t\to\infty} = \frac{1}{\delta}\sqrt{\frac{2}{b_o\Omega}}\int_0^t exp^{-b_1(t-s)}dW_s \tag{3.82}$$

Putting Eq. 3.82 back in Eq. 3.80, we get,

$$f'(0) = \frac{1}{2\sqrt{\left(k_1 a\frac{n^2}{\Omega} + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}}k_1 a\frac{n^2}{\Omega}\frac{1}{\delta}\sqrt{\frac{2}{b_o\Omega}}\int_0^t exp^{-b_1(t-s)}dW_s dW_t' \tag{3.83}$$

Wiener processes W_t and W_t' denote driving force behind intrinsic fluctuations in ξ' (or n_y) and n. Absence of correlation between W_t and W_t', leads to,

$$f'(0) = 0, \tag{3.84}$$

Hence, Eq. 3.74 can be reduced to,

$$dn = \left(k_1 a\frac{n^2}{\Omega} - k_2\frac{n^3}{\Omega} + k_3 b\Omega - k_4 n\right)dt + \sqrt{\left(k_1 a\frac{n^2}{\Omega} + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}dW_t' ... \\ ... + k_1 a\frac{n^2}{\Omega}\xi(t)dt, \tag{3.85}$$

or,

$$\frac{dn}{dt} = h(n) + g_2(n)\epsilon(t) + g_1(n)\xi(t), \tag{3.86}$$

Eq. 3.85 is an SDE including multiplicative white noise driven Wiener process $W_t' = \int \epsilon(t)dt$ and coloured noise $\xi(t)$. As mentioned in chapter 2, we can write an approximate FPE corresponding to Eq. 3.85 and extract a stationary state probability distribution. Following notations from chapter 2,

$$h(n) = \left(k_1 a\frac{n^2}{\Omega} - k_2\frac{n^3}{\Omega} + k_3 b\Omega - k_4 n\right), \tag{3.87}$$

$$g_1(n) = k_1 a\frac{n^2}{\Omega}, \tag{3.88}$$

$$g_2(n) = \sqrt{\left(k_1 a\frac{n^2}{\Omega} + k_2\frac{n^3}{\Omega} + k_3 b\Omega + k_4 n\right)}, \tag{3.89}$$

and

$$\alpha = \frac{1}{2}, \quad D = \frac{1}{b_o\Omega}, \quad \lambda = 0,$$

Hence, a stationary state distribution can be written as,

$$P_s(n) = \mathcal{N} \frac{G(\tau, n)}{\left(D g_1(n)^2 + \frac{1}{2} g_2(n)^2 \right)^{\frac{1}{2}}} exp\left(\int_0^n \frac{h(s) G(\tau, s)}{D g_1(s)^2 + \frac{1}{2} g_2(s)^2} ds \right), \quad (3.90)$$

where,

$$G(\tau, n) = 1 - \tau \left(h'(n) - \frac{g'_1(n)}{g_1(n)} h(n) \right), \quad (3.91)$$

As done previously, we can transform distribution of molecule number n to distribution of concentration x by change of variable. A comparison of stationary state distribution using UCNA with SSA is shown in Fig. 3.8 for reaction system R_1 for $\Omega = 30$ and 100. With increasing system size, agreement between UCNA and SSA distribution gets better. Similarly, results are also shown for R_2, R_3 and R_4, where external fluctuations of Y participate through other individual reactions of Schlögl model.

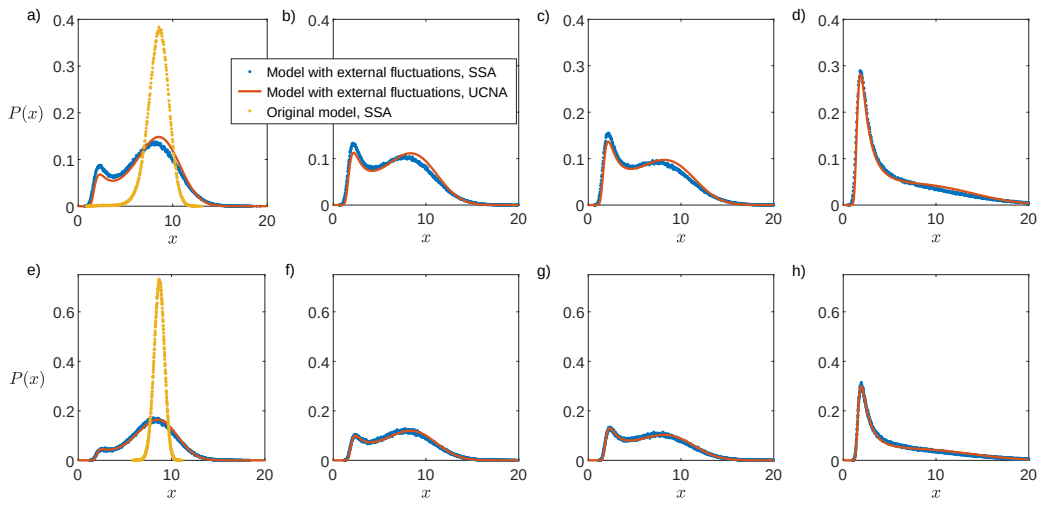

Figure 3.8: Agreement between stationary state concentration distribution derived using UCNA (orange line) with the one extracted from SSA simulations (blue dots) improves with increasing system size Ω, where X follows reaction set R_1 for two different system sizes $\Omega = 30$ (for a, b, c d) and 100 (for e, f, g, h). SSA distribution is extracted from single SSA trajectory. $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, b_1 = 1$. $b_o\Omega$ is 80 for (a) and (e), 50 for (b) and (f), 40 for (c) and (g), 10 for (d) and (h). Yellow dots indicate stationary state distribution of X in original Schlögl model.

Agreement between UCNA and SSA is very good for R_1 and R_4 (see Fig.

3.9), where external fluctuations are imposed on second (reaction set R_1) and first order reactions (reaction set R_4) respectively. For reaction set R_2, where external fluctuations influence propensity of a third order reaction, UCNA underestimates the lower mode of distribution (Fig. 3.10), though its distribution shows similar transition and doesn't lag behind in sensitivity in comparison to SSA. Similarly, for R_3, propensity fluctuations on a zero order reaction, UCNA predicts a wider distribution with smaller peak in comparison to SSA (Fig. 3.11), though both UCNA and SSA distribution remain almost unperturbed due to external fluctuations.

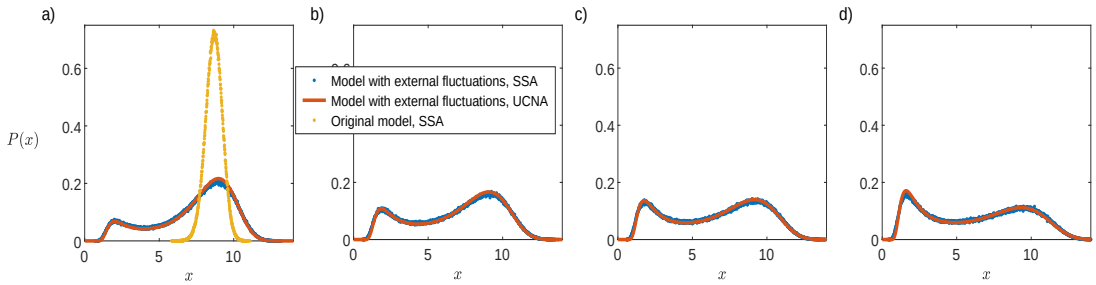

Figure 3.9: Good agreement between stationary state concentration distribution derived using UCNA (orange line) with the one extracted from SSA simulations (blue dots), where X follows reaction set R_4. SSA distribution is extracted from single SSA trajectory. $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, b_1 = 1, \Omega = 100$. $b_o\Omega$ is (a) 30, (b) 20, (c) 15 and (d) 10. Yellow dots indicate stationary state distribution of X in original Schlögl model.

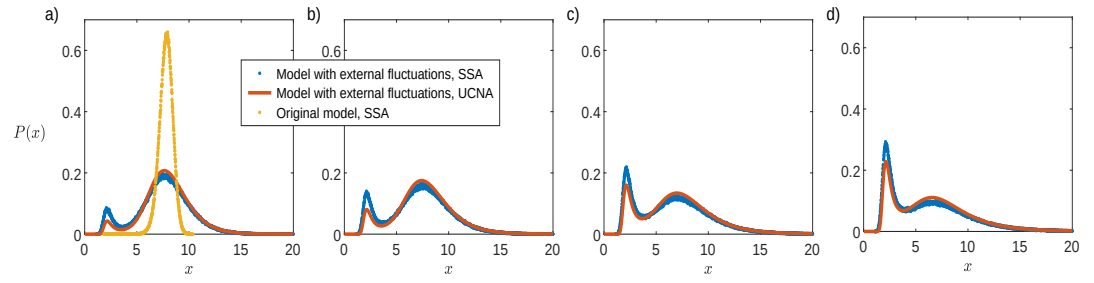

Figure 3.10: For reaction set R_2, where Y interacts with Schlögl Model through a third order reaction, stationary state distribution of X concentration derived using UCNA (orange line) overestimates the higher mode in comparison to the one extracted from SSA simulations (blue dots). SSA distribution is extracted from single SSA trajectory. $k_1 = 0.256, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, b_1 = 1, \Omega = 100$. $b_o\Omega$ is (a) 30, (b) 20, (c) 15 and (d) 10. Yellow dots indicate stationary state distribution of X in original Schlögl model.

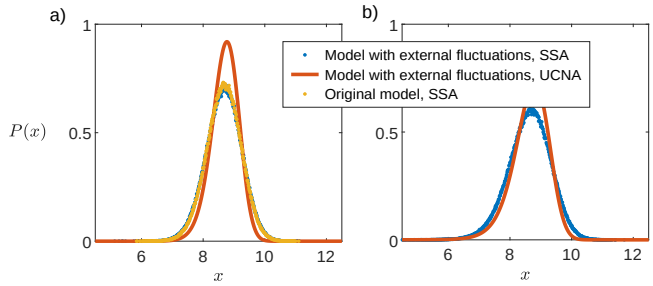

Figure 3.11: For reaction set R_3, where Y interacts with Schlögl Model through a zero order reaction, stationary state distribution of X concentration derived using UCNA (orange line) underestimates the distribution mode in comparison to the one extracted from SSA simulations (blue dots). SSA distribution is extracted from single SSA trajectory. $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, b_1 = 1, \Omega = 100$. Birth-rate $b_o\Omega$ is (a) 100, (b) 10. Yellow dots indicate stationary state distribution of X in original Schlögl model.

We can further elaborate on the sensitivity of Schlögl model towards external fluctuations through different reaction channels, using analytical expression for distribution extremum derived using UCNA. Fig. 3.12 presents a detailed picture of distribution sensitivity towards external fluctuations through different reaction channels. In agreement with previously shown transitions in Fig. 3.8-3.11, Fig. 3.12 demonstrates change in distribution extremum with change in intensity of external fluctuations and macroscopic reaction rates. For fixed model parameters, with increasing external fluctuations, higher mode decreases and lower mode increases, but more importantly, distribution minima (red dashed line) increases, hence indicating a shift of distribution weight towards lower mode (an increase in width of distribution around smaller peak). From figures 3.12, it is clear that different reaction channels have different sensitivities towards external fluctuations. In addition, reaction set R_1, where external fluctuations interact through a second order reactions, turns out to be most sensitive towards external fluctuations.

From Fig. 3.12, it is evident that presence of external fluctuations not only changes distribution weights around modes but also brings change to parameter regimes of feasible bimodality.

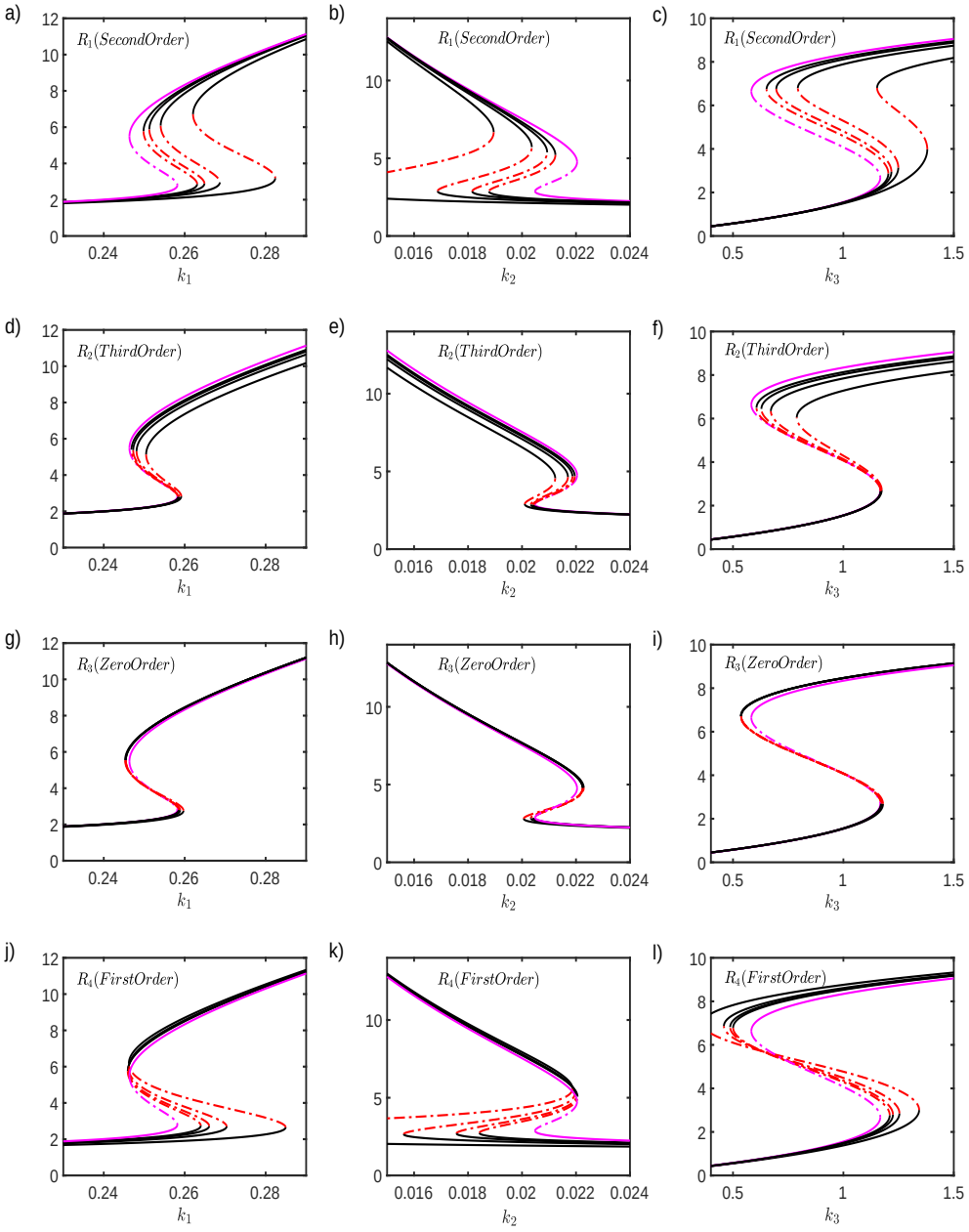

Figure 3.12: Change in distribution modes of Schl$\ddot{o}gl$ model corresponding to change in individual model parameters k_1, k_2 and k_3, and $\frac{b_1}{b_0\Omega}$, the change in intensity of external fluctuations Y, which interacts through different individual reactions (R_1, R_2, R_3, R_4). Analytical expression for distribution extremum is derived by putting first derivative of UCNA distribution to zero. In each figure, apart from the variable parameter, other parameters are fixed as: $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, a = 1, b = 1, b_1 = 1$ and $\Omega = 100$. $b_o\Omega$ varies from 100 to 25. Solid black lines indicate stable stationary states (modes), dashed red lines indicate unstable stationary states (distribution minima), pink lines indicate stationary states for original model.

3.3.1 Comparison of CFPE and UCNA against LNA

Even with the limited applications of CFPE and UCNA to single variable models, performance of both CFPE and UCNA turn out much more accurate in comparison to LNA. CFPE and UCNA captures the stochstic bimodality intuitively without imposing any further assumptions. On the other hand LNA fits normal distribution around deterministic stationary states, hence LNA can only recreate a bimodal distribution, when deterministic bistability coexist with stochastic bimodality. In fig. 3.13, we present a demonstration of how CFPE and UCNA perform better than LNA. It is to be mentioned that parameters are chosen to make sure that original Schl*ö*gl model from eq. 3.1 and 3.2 exhibit deterministic bistability (2 stable and 1 unstable deterministic stationary states) and unimodal mesoscopic stationary state probability distribution (1 extremum only). While, LNA have demonstrated sufficient accuracy in capturing bimodality in past few cases [83], here LNA fails to prove its utility, mostly due to highly non-linear nature of equations. While LNA predictions do give an unrefined idea of overall transitions from higher to lower mode transitions of distribution weight, it indeed fails in capturing parameter sensitivity along with analytically accurate distribution, as significant portion of concentration distribution lies below zero as shown in fig. 3.13(d).

As discussed previously, while CFPE and UCNA renders equations for stationary state probability distribution effortlessly, reconstructing bimodal distribution from LNA requires further amalgamation of two normal distributions. Steps to reconstruct an approximated bimodal distribution using LNA are as following:

(i) Derive two separate covariance matrices for both deterministic stable stationary states.

(ii) Construct two individual normal distributions $(\mathcal{N}_1, \mathcal{N}_2)$, using individual deterministic stable stationary states as mean and variance of X from covariance matrix corresponding to each stable stationary state.

(iii) Assume, both distributions merge at unstable stationary state.

(iv) For normal distribution with larger mean, calculate cumulative probability P_1 from $-\infty$ to unstable stationary state.

(v) For normal distribution with smaller mean, calculate cumulative probability P_2 from unstable stationary state to ∞.

(vi) Construct a weighted distribution from the individual normal distributions as,

$$P_{LNA} = \frac{P_1}{P_1 + P_2}\mathcal{N}_1 + \frac{P_2}{P_1 + P_2}\mathcal{N}_2,$$

where, $\mathcal{N}_1$ corresponds to normal distribution with larger mean and $\mathcal{N}_2$ corresponds to normal distribution with smaller mean.

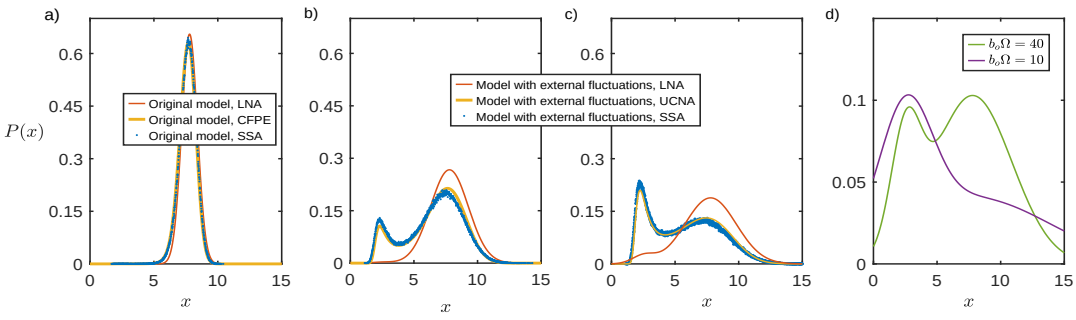

Figure 3.13: CFPE and UCNA naturally captures mesoscopic bimodality and show accurate sensitivity toward parameter changes, while LNA remains severely far behind with respect to rendering an accurate distribution and parameter sensitivity. For original Schlögl model in (a) and model with external fluctuations (reaction set R_1) in (b) and (c) stationary state distribution of X concentration derived using CFPE and UCNA (yellow line) in comparison with stationary state distribution reconstructed using LNA (red line). While LNA do capture transition in modes as shown in (d) but performs poorly with regards to overall accuracy of probability distribution as well as parameter sensitivity. SSA distribution (blue dots) are extracted from single SSA trajectory. $k_1 = 0.255, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, b_1 = 1, \Omega = 100$. For model with external fluctuations, birth-rate $b_o\Omega$ is (b) 200, (c) 100 and (d) 40 and 10 respectively.

3.3.2 Limitations of UCNA as an analytical tool

While we showed utility of UCNA as an approximate analytical tool to capture change in mesoscopic bimodality due to external fluctuations, it is also important to point out a few limitations. First which arises is that, our analysis is limited only to sufficiently large system sizes and small external fluctuation intensity [57]. This is due to 1) approximation of discrete birth-death process to a continuous diffusion process which leads to differences between Poisson and Normal distributions as shown in Fig. 3.6, 2) use of CFPE as a starting point and 3) inherent approximation in UCNA itself, which limits its applicability for small noise limit . Secondly, similar to possible discrepancies between detailed balance recursive solution and approximated distribution derived using CFPE of Schlögl model, in the case of extremely small external fluctuations ($\frac{b_o\Omega}{b_1} \gg 1$), the use of UCNA leads to an inherent bias towards higher mode of distribution. This becomes more evident if the original distribution consists of considerable bimodality as shown in Fig. 3.14.

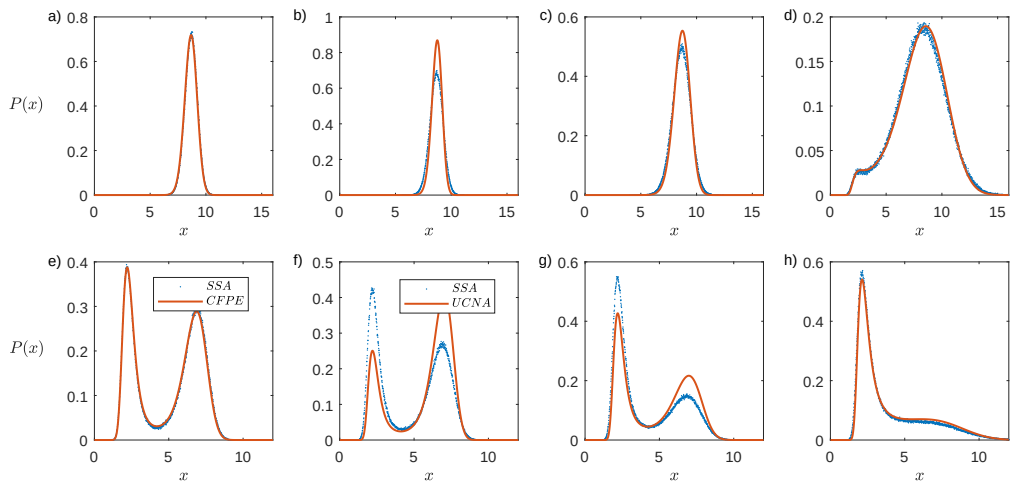

Figure 3.14: UCNA predictions deviate from exact SSA distribution for reaction set R_1, especially when Schlögl model has apparent bimodalty itself. (a) and (e) show SSA and CFPE distributions for original model without any external fluctuations. For (a), (b), (c) and (d) $k_1 = 0.263$, for (e), (f), (g) and (h) $k_1 = 0.25$. $b_o\Omega$ varies as 10^4 for (b) and (g), 10^3 for (c) and (h), 10^2 for (d) and (i). Other parameters are $k_2 = 0.0188, k_3 = 1.203, k_4 = 1, a = 1, b = 1, b_1 = 1$ and $\Omega = 100$. Notice the different scales on x-axis.

Another more severe limitation of UCNA comes in the form of breakdown of distribution shown in Eq. 3.90, when $G(\tau, n)$ (Eq. 3.91) becomes negative. As shown in Eq. 3.91, model parameters for which $\tau\left(h'(n) - \frac{g_1'(n)}{g_1(n)}h(n)\right) > 1$, will lead to nonsensical negative probability densities. An example of this is illustrated in Fig 3.15, where breakdown of probability distribution is shown due to slower time scales of external fluctuations.

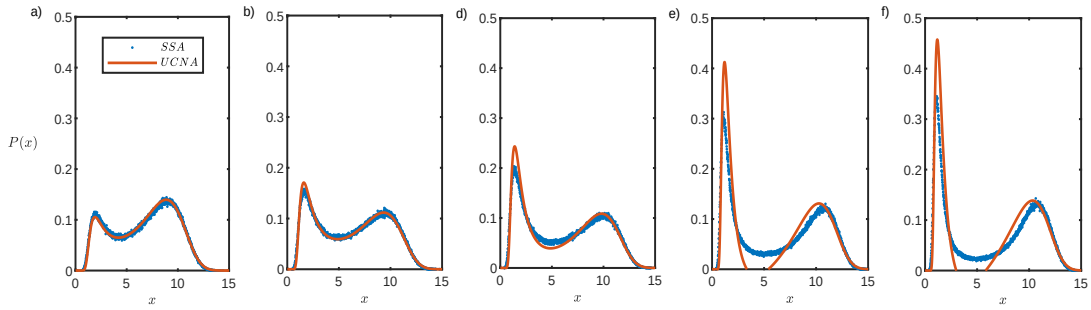

Figure 3.15: UCNA breaks down and renders spurious results for reaction set R_4, with external fluctuations at slower time scales. Times scale τ of external fluctuation is (a) 0.5, (b) 1, (c) 2, (d) 10 and (e) 20 with mean $\langle Y \rangle_s$ fixed to 10. Other parameters are $k_1 = 0.263, k_2 = 0.0188, k_3 = 1.203, k_4 = 1, a = 1, b = 1$ and $\Omega = 100$.

3.4 Conclusion

Schlögl model is the simplest reaction network which exhibits bistability [84]. Various significant biological mechanisms can be mapped to positive auto-regulation mechanisms, similar to Schlögl model [60]. Analysis of bistability and bimodality have lead to extensive novel analysis in computational biology. A similar pertinent example includes Schlögl model as a minimal model to study effects of burst-noise [69]. Understanding stochastic kinetics of Schlögl model provides a comprehensive platform to study mesoscopic effects of interactions between different biological mechanisms [60] and external sources of fluctuations. Our analysis further characterizes bimodal transitions due to network topology of Schlögl model, in communication with its surroundings and establishes UCNA as an

analytical tool to capture these transitions. Analysis on effects of external fluctuations through different reaction channels presents a simplistic approach toward capturing complexities of intracellular processes. We have focused on simple birth-death mechanism as a source of external fluctuations, which are widely used to model upstream transcription as well as dynamics of various catalytic enzymes. Hence, in the next chapter, we will discuss effects of similar dynamic external sources on stochastic responses of various gene expression mechanisms.

Chapter 4

EXTERNAL FLUCTUATIONS IN GENE EXPRESSION

4.1 Introduction

In this chapter, we further extend applications of UCNA to gene expression mechanisms. We will discuss two popular gene expression mechanisms used in systems biology for quantitative understanding of stochasticity in protein translation. Later on, we present comparison of SSA simulations with UCNA predictions, hence proving utility of UCNA as an analytical approximation of CME in context of systems biology. Using our analysis, we demonstrate mesoscopic transitions driven by fluctuating mRNAs as well as fluctuations in enzymatic machineries, which are often ignored for the sake of simplification and solvability during analytical treatment.

4.2 Gene Expression Mechanisms in Cell Biology

We mainly focus on two popular primary mechanisms which provide schematic understanding of central dogma and its topological regulation. To begin with, as shown in Eq. 4.1, the first mechanism is gene expression from

a constitutive promoter [14, 66].

$$\begin{aligned} Gene &\xrightarrow{k_1} mRNA + Gene \\ mRNA &\xrightarrow{r_1} \Phi \\ mRNA &\xrightarrow{k_2} Protein + mRNA \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{4.1}$$

Eq. 4.1 is a simplistic scheme of central dogma consisting of transcription from a single gene, translation from mRNA and degradation of mRNA and protein. Here, mRNA acts as an intermediate template for protein synthesis. This mechanism is also known as 2-stage model of gene expression. An approximate analytical solution of corresponding CME is provided in [14] which fits a negative binomial distribution for protein while time scale of mRNA degradation $(\frac{1}{r_1})$ is much faster than time scale of protein degradation $(\frac{1}{r_2})$.

Another mechanism as part of our analysis on effects of external fluctuations, is self-regulatory non-linear feedback [15], as shown in Eq. 4.2.

$$\begin{aligned} Gene + 2\,Protein &\underset{p_2}{\overset{p_1}{\rightleftharpoons}} Gene^* \\ Gene &\xrightarrow{k_0} Protein + Gene \\ Gene^* &\xrightarrow{k_1} Protein + Gene^* \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{4.2}$$

Non-linear feedback can be regarded as cooperativity in biological context. Depending upon magnitude of basal (from G) and activated (from G^*) translation rates (k_0 and k_1), we can define up and down regulation of translation which is also known as positive and negative feedback.

As discussed in chapter 1, mechanism shown in Eq. 4.2 is popular in systems biology even when it skips intermediate transcription. Generally, these mechanisms are based on quasi-steady state assumption [51], which is applied to mRNA due to its faster dynamics time scale in comparison to that of protein. While this time scale difference is true for a large number of gene expression pathways, advent of high-throughput technology in synthetic biology brings in feasibility to tinker mRNA stability by changing secondary structures as well as interaction with degradation enzyme [85–89]. This makes it necessary to consider transcription dynamics while evaluating stochastic distributions of expressing proteins. Similarly, in higher order organisms, mRNAs with much slower degradation time scales are not so rare to find.

In our quantitative analysis we will focus on analysing effects of fluctuating mRNA [90] and proteolytic enzyme on protein distribution in gene expression mechanisms from Eq. 4.2. We will begin with implementing SSA and UCNA as numerical and analytical tools for stochastic analysis of mechanism in Eq. 4.1, as a test case to understand transcription and proteolytic enzyme as sources of external fluctuations.

4.3 External Fluctuations in Expression from a Constitutive Promoter

4.3.1 Transcription as upstream external fluctuation

As we discussed application of quasi-steady state assumption to remove transcription from gene expression mechanism, a simple example can be demonstrated using mechanism from Eq. 4.1. We can reduce it to a simple

birth-death process,

$$\begin{aligned} Gene &\xrightarrow{k_{eff}} Protein + Gene \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{4.3}$$

where, $k_{eff} = \frac{k_1}{r_1} k_2$ leads to same deterministic stationary state for protein, as in mechanism from Eq. 4.1. To bring in stochastic fluctuations of mRNA, we can transform mechanism in Eq. 4.3 as,

$$\begin{aligned} Gene &\xrightarrow{\frac{1}{[mRNA]_s} k_{eff}[mRNA]} Protein + Gene \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{4.4}$$

where, mRNA follows a simple birth-death mechanism, which, similar to our previous analysis in chapter 3, can be approximated with an OU process. This transformation is exactly the same as what we did previously with Schlögl model and transforming constitutive gene expression to a one-variable model with an upstream external noise source. We can write an approximate stationary state probability distribution using UCNA, similar to what we discussed in chapter 3. In the context of UCNA derivation, discussed in chapter 2, we can write an OU process driven SDE for reaction system in eq. 4.4 as,

$$\dot{x} = h(x) + g_1(x)\epsilon(t) + g_2(x)\Gamma(t), \tag{4.5}$$

where,

$$\begin{aligned} h(n) &= k_{eff} - r_2 n, \\ g_1(n) &= k_{eff}, \\ g_2(n) &= \sqrt{k_{eff} + r_2 n} \end{aligned}$$

$$\text{with} \quad D = \frac{1}{k_1}, \quad \tau = \frac{1}{r_1} \text{ and } \lambda = 0$$

where, n is number of protein molecules and Ω is system size. Using terms from Eq. 4.5 and bringing UCNA into action, we can derive an analytical expression for stationary state distribution. In Fig. 4.1, we show a comparison of protein distribution computed using both SSA and UCNA. Model parameters are defined the same way as in [14] where $\gamma = \frac{r_1}{r_2}$, $a = \frac{k_1}{r_2}$ and $b = \frac{k_2}{r_2}$, where γ denotes the ratio of time scales of protein and mRNA dynamics.

From Fig. 4.1, it is clear that, though UCNA applications are limited for gene expression with sufficiently large expression sizes, yet it proves utility by accurate estimation of stationary state protein distribution even for small γ values, which is not the case for solution derived in [14]. It is interesting to note that while, UCNA outperforms the solution given in [14] for $\gamma = 1$ but UCNA solution deteriorates for $\gamma = 10$. Similar to previous chapter, this fall in UCNA accuracy is due to smaller stationary state mean of intermediate template mRNA ($\langle n_{mRNA} \rangle_s = 2$, for $\gamma = 10$), which eventually affects normal approximation of Poisson distribution. Hence, UCNA can provide significant analytical insights for gene expression systems where mRNA shows higher stability and its half-life is of a similar time scale as that of a protein.

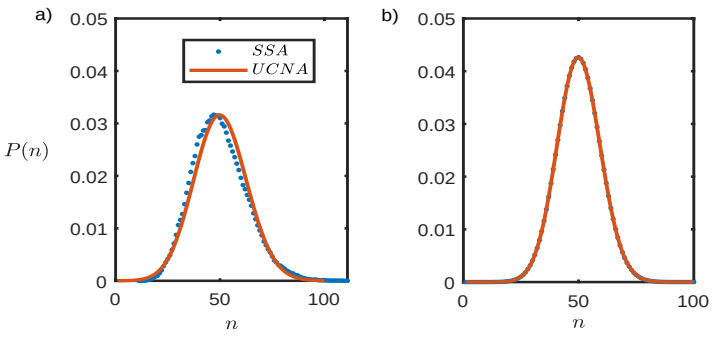

Figure 4.1: For a 2-stage, constitutive promoter mechanism, superiority of UCNA is evident at large protein numbers, in comparison to previously derived exact solution, as UCNA derived stationary state protein distributions are in good agreement with SSA, especially for smaller γ. UCNA approximation treats upstream transcription as external fluctuations. (a) $\gamma = 10$ and (b) $\gamma = 1$. Other parameters are $a = 20$, $b = 2.5$, $r_2 = 0.0005$ and $\Omega = 1$, where $\gamma = \frac{r_1}{r_2}$, $a = \frac{k_1}{r_2}$ and $b = \frac{k_2}{r_2}$.

4.3.2 Proteolytic enzyme as downstream external fluctuation

We can also focus on downstream noise in translation kinetics due to fluctuations in proteolytic enzymes, instead of the usual upstream transcription noise as discussed previously. Proteolytic enzymes which are responsible for degradation of unnecessary proteins, considered to have their own stochastic dynamics, following simple birth-death translation. Hence, a simple schematic of this mechanism can be written as,

$$\begin{aligned} Gene &\xrightarrow{k_{eff}} Protein + Gene \\ Protein &\xrightarrow{\frac{1}{[E]_s} r_2 E} \Phi \end{aligned} \tag{4.6}$$

where, E denotes the downstream proteolytic enzyme and $k_{eff} = \frac{k_1}{r_1} k_2$. Similar to transcription (or birth-death translation 4.3), proteolytic enzyme E also follows a simple birth-death kinetics with macroscopic birth and death rates as $k_3 = (\frac{k_1}{\Omega})$ and $r_3 (= r_1)$ respectively. In this case, we can recall UCNA

with following changes and write in context of eq. 4.5,

$$\begin{aligned} h(n) &= k_{eff} - r_2 n, \\ g_1(n) &= -r_2 n, \\ g_2(n) &= \sqrt{k_{eff} + r_2 n} \end{aligned} \tag{4.7}$$

$$\text{with} \quad D = \frac{1}{k_3 \Omega}, \quad \tau = \frac{1}{r_3} \text{ and } \lambda = 0$$

Fig. 4.2 demonstrates right skewness in stationary state protein distribution due to downstream fluctuations, which was previously absent in Fig. 4.1.

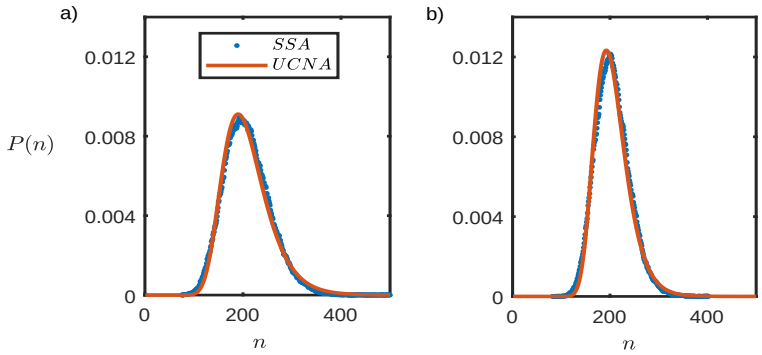

Figure 4.2: When protein follows simple birth-death translation (no upstream transcription) and downstream proteolytic fluctuations, UCNA derived stationary state protein distributions at large protein numbers are in good agreement with SSA. UCNA approximation treats downstream proteolytic enzyme kinetics as external fluctuations. (a) $\gamma = 10$ and (b) $\gamma = 1$. Other parameters are $a = 20$, $b = 10$, $r_2 = 0.0005$, and $\Omega = 1$.

Applications of UCNA on simple 2-stage model of gene expression as well as on translation with fluctuating proteolysis shows good agreement with SSA simulations, hence as a next step, we further its applications to gene expression mechanism with a non-linear feedback.

4.4 External Fluctuations in Expression including a 2-state Promoter with Non-Linear Feedback

In this section, we have provided analysis on effects of both up and downstream external fluctuations on gene expression mechanism with 2-state random switching promoter with a non-linear feedback (Eq. 4.2), in its different parametric configurations which include fast and slow switching between basal and activated states. While a more accurate solution, independent of system size Ω parameters have recently been published [91], analytical simplicity of our analysis makes it distinct.

4.4.1 Quasi-steady state approximation for a fast switching promoter

For switching rates between basal and activated states (p_1, p_2) much higher in comparison to protein dynamics time scales, we can simplify this mechanism by assuming a quasi-steady state of gene states. To begin with, we can start with writing deterministic REs for basal gene state concentration G and protein concentration P,

$$\frac{dG}{dt} = -p_1 G P^2 + p_2(\frac{1}{\Omega} - G), \tag{4.8}$$

$$\frac{dP}{dt} = -\frac{1}{2}p_1 G P^2 + \frac{1}{2}p_2(\frac{1}{\Omega} - G) + k_o G + k_1(\frac{1}{\Omega} - G) - r_2 P, \tag{4.9}$$

where $\frac{1}{\Omega}$ refers to concentration of single gene copy in both basal and activated state combined. Now, for fast switching rates, we can assume a

quasi-steady state between states, which using Eq. 4.8 gives,

$$G = \frac{1}{\Omega}\frac{p_2}{p_1 P^2 + p_2}, \tag{4.10}$$

Hence, Eq. 4.9 can be reduced to,

$$\frac{dP}{dt} = \frac{1}{\Omega}\frac{k_o p_2 + k_1 p_1 P^2}{p_2 + p_1 P^2} - r_2 P, \tag{4.11}$$

or,

$$\frac{dn}{dt} = \frac{k_o p_2 + k_1 p_1 \frac{n^2}{\Omega^2}}{p_2 + p_1 \frac{n^2}{\Omega^2}} - r_2 n, \tag{4.12}$$

where, n is number of molecules of protein. Eq. 4.12 represents one-variable dynamics for which analytical distribution using CLE/CFPE can be derived with much ease [59]. As discussed in chapter 2, CLE corresponding to reaction system from eq. 4.2 can be written as,

$$dn = \left(\frac{k_o p_2 + k_1 p_1 \frac{n^2}{\Omega^2}}{p_2 + p_1 \frac{n^2}{\Omega^2}} - r_2 n\right)dt + \sqrt{\left(\frac{k_o p_2 + k_1 p_1 \frac{n^2}{\Omega^2}}{p_2 + p_1 \frac{n^2}{\Omega^2}} + r_2 n\right)}dW_t, \tag{4.13}$$

Stationary state distribution of protein molecules n from eq. 4.13, is shown in fig. 4.3. It captures the transitions in stationary state bimodal distribution with change in protein degradation rate r_2, during positive feedback ($k_1 > k_0$).

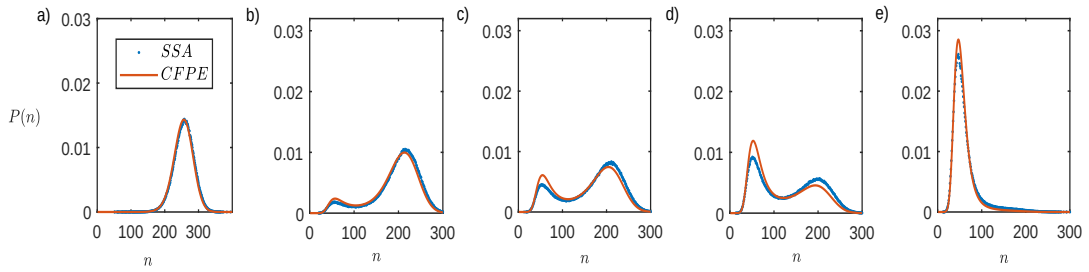

Figure 4.3: For protein translation through a positive feedback and fast 2-state random switching promoter, CFPE approximates stationary state protein distribution in good agreement with SSA at large protein numbers. For both CFPE and SSA, distribution bimodality changes with change in protein degradation rate r_2. r_2 = (a) 1.18 (b) 1.24, (c) 1.25, (d) 1.26 and (e) 1.3. Other parameters are $k_0 = 0.2\Omega$, $k_1 = 2.32\Omega$, $p_1 = p_2 = 1000$ and $\Omega = 200$.

4.4.2 Fast switching promoter with upstream transcription and downstream proteolytic enzyme fluctuations

In continuity with our focus on analysing effects of external components on the stochastic behaviour of innate mechanism, we bring upstream transcription and downstream proteolytic enzyme fluctuations into the picture, individually. To begin with, mechanism shown in Eq. 4.2 can be further extended to include transcription as following,

$$\begin{aligned}
Gene + 2\,Protein &\underset{p_2}{\overset{p_1}{\rightleftharpoons}} Gene^* \\
Gene &\xrightarrow{k_0'[mRNA]} Protein + Gene \\
Gene^* &\xrightarrow{k_1'[mRNA]} Protein + Gene^* \\
\Phi &\xrightarrow{k_{tran}} mRNA \\
mRNA &\xrightarrow{r_1} \Phi \\
Protein &\xrightarrow{r_2} \Phi
\end{aligned} \tag{4.14}$$

where, k_{trans} is defined rate of transcription. Setting $k_0' = \frac{r_1}{k_{trans}} k_0$ and $k_1' = \frac{r_1}{k_{trans}} k_1$ keeps the deterministic stationary states same as deterministic stationary states for mechanism in Eq. 4.2. As discussed in chapter 3, LNA provides an adequate OU process approximation for stochastic kinetics of

mRNA molecules as,

$$dn_{mRNA} = -r_1(n_y - \frac{k_{trans}}{r_1})dt + \sqrt{2k_{trans}\Omega}dW_t, \tag{4.15}$$

Hence, in the context of UCNA and transcription as source of coloured noise, we can transform this mechanism as,

$$\begin{aligned} Gene + 2\,Protein &\underset{p_2}{\overset{p_1}{\rightleftharpoons}} Gene^* \\ Gene &\xrightarrow{\frac{1}{[mRNA]_s}k_0[mRNA]} Protein + Gene \\ Gene^* &\xrightarrow{\frac{1}{[mRNA]_s}k_1[mRNA]} Protein + Gene^* \\ Protein &\xrightarrow{r_2} \Phi \end{aligned} \tag{4.16}$$

where $[mRNA]_s = \frac{k_{trans}}{r_1}$. Now, bringing all pieces together including drift-diffusion terms of CFPE and OU process approximation of mRNA dynamics, we can apply UCNA as discussed for previous case in eq. 4.5 with following terms and write stationary state distribution for proteins n.

$$\begin{aligned} h(n) &= \frac{k_0K\Omega^2 + k_1n^2}{K\Omega^2 + n^2} - r_2n, \\ g_1(n) &= \frac{k_0K\Omega^2 + k_1n^2}{K\Omega^2 + n^2}, \\ g_2(n) &= \sqrt{\frac{k_0K\Omega^2 + k_1n^2}{K\Omega^2 + n^2} + r_2n} \end{aligned} \tag{4.17}$$

$$\text{with } K = \frac{p_2}{p_1}, \quad D = \frac{1}{k_{trans}\Omega}, \quad \tau = \frac{1}{r_1} \text{ and } \lambda = 0$$

Fig. 4.4 and 4.5 show comparison of UCNA with SSA simulations for increasing intensity of transcriptional fluctuations during positive and negative feedbacks. From mechanistic point of view, role of transcription and mRNA as an intermediate is evident. In the case of positive feedback

$(k_1 > k_0)$, with increasing intensity of mRNA fluctuations, distribution weight switches from higher mode to lower mode, indicating a widely observed phenotypic switching which carries the potential to play an important role in synthetic biological circuits.

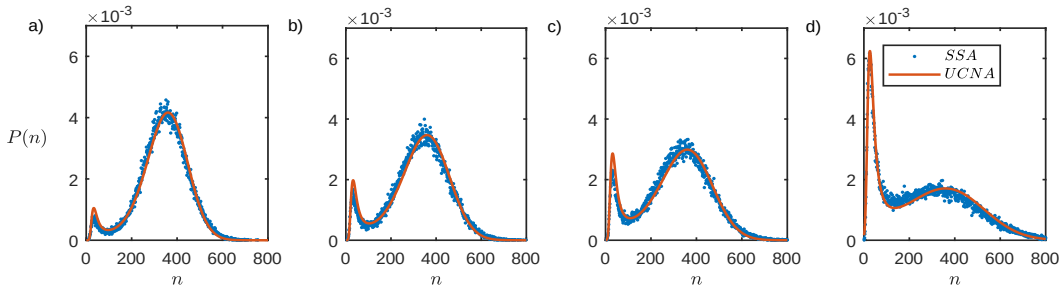

Figure 4.4: For protein translation through a positive feedback and 2-state fast random switching promoter, stationary state distribution switches from higher to lower mode with increase in upstream transcriptional noise. For large protein numbers, UCNA is in agreement with SSA. $\frac{\tau}{D}$ = (a) 20 (b) 15, (c) 12 and (d) 5. Other parameters are $r_1 = r_2 = 1$, $k_0 = 0.12\Omega$, $k_1 = 2.32\Omega$, $p_1 = p_2 = 1000$ and $\Omega = 200$.

Similarly, for negative feedback $(k_0 > k_1)$, increase in transcription fluctuations leads to wider stationary state distributions, indicating higher phenotypic variations.

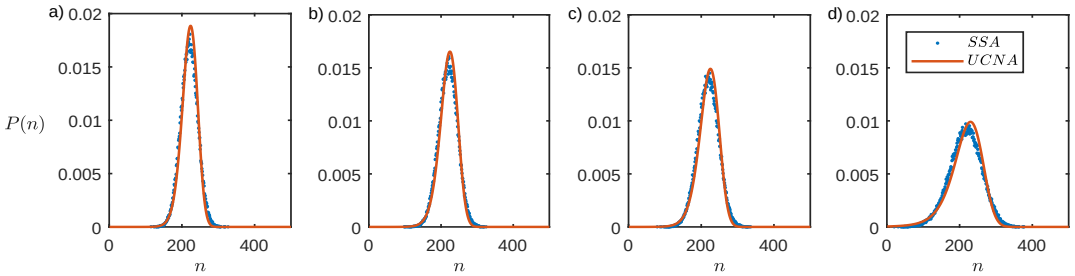

Figure 4.5: For protein translation through a negative feedback and 2-state fast random switching promoter, stationary state distribution variance increases with increase in upstream transcriptional noise. For large protein numbers, UCNA is in agreement with SSA. $\frac{\tau}{D}$ = (a) 20 (b) 15, (c) 12 and (d) 5. Other parameters are $r_1 = r_2 = 1$, $k_0 = 2.32\Omega$, $k_1 = 0.12\Omega$, $p_1 = p_2 = 1000$ and $\Omega = 200$.

Following a similar approach, we can further test the impact of downstream stochasticity due to fluctuations in proteolytic enzyme with birth and death rates k_3 and r_3 respectively. In this case, similar to the mechanism shown in Eq. 4.5, we can include a proteolytic enzyme E facilitating degradation of

protein in the mechanism from Eq. 4.2.

$$\begin{aligned} Gene + 2\,Protein &\underset{p_2}{\overset{p_1}{\rightleftharpoons}} Gene^* \\ Gene &\xrightarrow{k_0} Protein + Gene \\ Gene^* &\xrightarrow{k_1} Protein + Gene^* \\ Protein &\xrightarrow{\frac{1}{[E]_s} r_2 E} \Phi \end{aligned} \tag{4.18}$$

UCNA can be implemented with following terms,

$$\begin{aligned} h(n) &= \frac{k_0 K\Omega^2 + k_1 n^2}{K\Omega^2 + n^2} - r_2 n, \\ g_1(n) &= -r_2 n, \\ g_2(n) &= \sqrt{\frac{k_0 K\Omega^2 + k_1 n^2}{K\Omega^2 + n^2} + r_2 n} \end{aligned} \tag{4.19}$$

$$\text{with } K = \frac{p_2}{p_1}, \quad D = \frac{1}{k_3\Omega}, \quad \tau = \frac{1}{r_3} \text{ and } \lambda = 0$$

As shown in Fig. 4.6 and 4.7, similar to upstream transcription, noise in proteolytic enzyme leads to bimodality transitions during positive feedback and wider distributions during negative feedback.

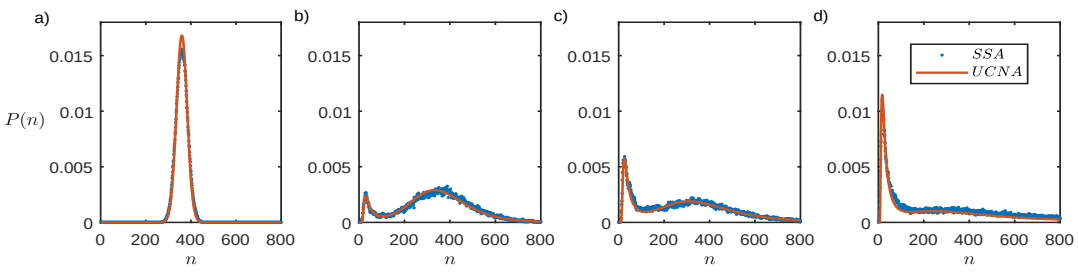

Figure 4.6: For protein translation through a positive feedback and 2-state fast random switching promoter, stationary state distribution switches from higher to lower mode with increase in downstream fluctuations in proteolytic enzyme. For large protein numbers, UCNA is in agreement with SSA. under influence of downstream proteolytic enzyme fluctuations. $\frac{\tau}{D}$ = (a) 20 (b) 15, (c) 12 and (d) 5. Other parameter are $r_2 = r_3 = 1$, $k_0 = 0.12\Omega$, $k_1 = 2.32\Omega$, $p_1 = p_2 = 1000$ and $\Omega = 200$.

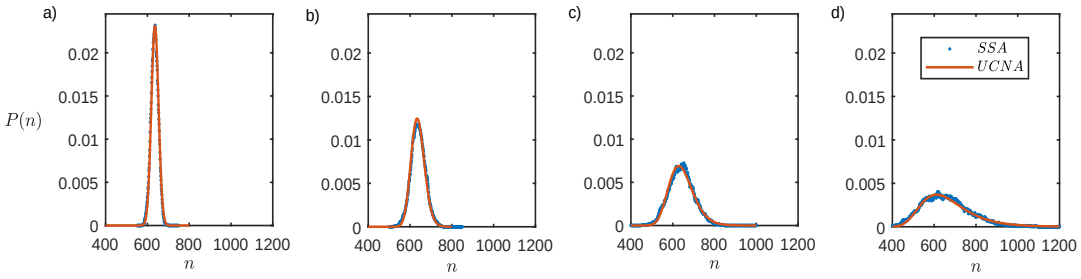

Figure 4.7: For protein translation through a negative feedback and 2-state fast random switching promoter, stationary state distribution variance increases with increase in downstream fluctuations in proteolytic enzyme. For large protein numbers, UCNA is in agreement with SSA. $\frac{\tau}{D}$ = (a) 20 (b) 15, (c) 12 and (d) 5. Other parameter are $r_2 = r_3 = 1$, $k_0 = 2.32\Omega$, $k_1 = 0.12\Omega$, $p_1 = p_2 = 1000$ and $\Omega = 200$.

In summary, combination of quasi-steady state approximation and UCNA works remarkably well to capture changes in protein distributions led by various external components.

4.5 Slow Switching Promoter: Model Reduction based on Protein Abundance

Good agreement between our analytical approximation using UCNA and SSA simulations, bolstered our confidence to further extend this approach to slow switching promoter systems where promoter switching rates p_1 and p_2 are of similar order as protein degradation rate. It has been shown that slow, random promoter switching itself leads to bimodality in gene expression, without requirements of non-linear feedback [14, 92]. In this case, quasi-steady state will not be a good approximation as it turns out previously in fast switching. To deal with this situation, we pursue another approach which is based on treating gene expression from basal and activated gene states, as it is from two separate individual constitutive promoters, and creating a final stationary state protein distribution as a mixture of individual stationary state distributions [83]. This mixture of two

separate distributions is based on the assumption that protein abundance is significantly higher than number of genes/promoters (which is 1). With this assumption, we can transform the original mechanism (Eq. 4.2) as following,

Expression from basal state:

$$\begin{aligned} Gene &\xrightarrow{k_0} Protein + Gene \\ Protein &\xrightarrow{r_2} \Phi \end{aligned}$$

Expression from activated state:

$$\begin{aligned} Gene^* &\xrightarrow{k_1} Protein + Gene^* \\ Protein &\xrightarrow{r_2} \Phi \end{aligned}$$

Gene switching:

$$Gene \underset{p_2}{\overset{p_1[Protein]^2}{\rightleftharpoons}} Gene^* \tag{4.20}$$

This simplification helps us in implementing UCNA as we have already shown how UCNA works remarkably well for capturing effects of both upstream and downstream external fluctuations in gene expressions.

Before we jump to analysing effects of external fluctuations, it is convenient to verify the gene state switching simplification. We start with CFPE to approximate stationary state distribution for both basal and activated gene states separately. In accordance with CFPE formulation explained previously,

we can write drift and diffusion terms as,

For basal gene state:

$$\begin{aligned} A_1(n) &= k_0 - r_2 n, \\ B_1(n) &= \sqrt{k_0 + r_2 n}, \end{aligned}$$

For activated gene state:

$$\begin{aligned} A_2(n) &= k_1 - r_2 n, \\ B_2(n) &= \sqrt{k_1 + r_2 n}, \end{aligned}$$

Let's say $P_G(n)$ is stationary state distribution for protein stationary state corresponding to basal state and $P_{G^*}(n)$ is protein stationary state distribution for activated state. Then, we can compute a final stationary state distribution $P(n)$ as a weighted mixture of $P_G(n)$ and $P_{G^*}(n)$.

$$P(n) = Prob_{basal} P_G(n) + (1 - Prob_{basal}) P_{G^*}(n), \tag{4.21}$$

where $Prob_{basal}$ can be defined as the probability of gene being in basal state, or fraction of per unit time gene spent in basal state, and calculated following approximated gene switching kinetics based on protein abundance. Hence, probability of gene being in basal state can be considered equal to normalized propensity of reaction $Gene^* \rightarrow Gene$, $Pr_{G^* \rightarrow G}$ which can be written as,

$$Prob_{basal} = \frac{Pr_{G^* \rightarrow G}}{Pr_{G^* \rightarrow G} + Pr_{G \rightarrow G^*}}, \tag{4.22}$$

where,

$$Pr_{G \rightarrow G^*} = p_1 \frac{\left(\int_0^\infty n P_G(n) dn\right)^2}{\Omega^2} \text{ and } Pr_{G^* \rightarrow G} = p_2$$

For both positive and negative feedbacks, comparison between the above analytical approximation and SSA simulations is shown in Fig. 4.8, where our analytical approximation is in good agreement during both positive and negative feedback parameter regimes.

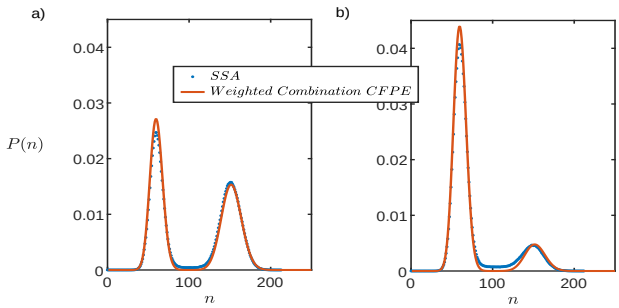

Figure 4.8: For protein translation through 2-state slow random switching promoter, stationary state distribution estimated as weighted combination of two CFPE distributions is in good agreement with one extracted using SSA. (a) $k_0 = 0.15\Omega$, $k_1 = 0.38\Omega$ (positive feedback), (b) $k_0 = 0.38\Omega$, $k_1 = 0.15\Omega$ (negative feedback). Other parameters are $r_2 = 0.5$, $k_0 = 0.15\Omega$, $k_1 = 0.38\Omega$, $p_1 = 0.1$, $p_2 = 0.01$, $\tau = 0.5$ and $\Omega = 200$.

4.5.1 Slow switching promoter with upstream transcription and downstream proteolytic enzyme fluctuations

To incorporate transcription noise from the perspective of analytical approximation, we modified simplified mechanism from Eq. 4.20 as following,

Expression from basal state:

$$Gene \xrightarrow{\frac{1}{[mRNA]_s} k_0 [mRNA]} Protein + Gene$$

$$Protein \xrightarrow{r_2} \Phi$$

Expression from activated state:

$$Gene^* \xrightarrow{\frac{1}{[mRNA]_s} k_1 [mRNA]} Protein + Gene^*$$

$$Protein \xrightarrow{r_2} \Phi$$

Gene switching:

$$Gene \quad \underset{p_2}{\overset{p_1[Protein]^2}{\rightleftharpoons}} \quad Gene^* \tag{4.23}$$

where, mRNA follows exactly the same kinetics as explained previously during fast switching mechanism (Eq. 4.14 and 4.15). Following the similar methodology as described in previous section using CFPE and weighted combination of distributions, we can derive analytical description for the above mechanism. The only difference is approximating mRNA dynamics with an OU process and implementing UCNA. Hence, UCNA can be used with following terms:

For protein n stationary state distribution $P_G(n)$ from gene basal state G:

$$\begin{aligned} h(n) &= k_0 - r_2 n, \\ g_1(n) &= k_0, \\ g_2(n) &= \sqrt{k_0 + r_2 n} \end{aligned} \tag{4.24}$$

$$\text{with} \quad D = \frac{1}{k_{trans}\Omega}, \quad \tau = \frac{1}{r_1} \text{ and } \lambda = 0$$

and for protein n stationary state distribution $P_{G^*}(n)$ from gene activated state G^*:

$$\begin{aligned} h(n) &= k_1 - r_2 n, \\ g_1(n) &= k_1, \\ g_2(n) &= \sqrt{k_1 + r_2 n} \end{aligned} \tag{4.25}$$

$$\text{with} \quad D = \frac{1}{k_{trans}\Omega}, \quad \tau = \frac{1}{r_1} \text{ and } \lambda = 0$$

Hence, following Eq. 4.22 and using $P_G(n)$ and $P_{G^*}(n)$, we can derive a weighted distribution $P(n)$ which captures stationary state probability distribution of protein n. In Fig. 4.9 and 4.10, we depict comparison between stationary state protein distribution derived using SSA simulations of mechanism from Eq. 4.23 and the previously described stationary state protein distribution derived using UCNA and weighted combination of distribution. It is clearly evident that similar to previous cases, UCNA distribution in both positive and negative feedback cases is in good agreement with SSA distribution. In this case, the final distribution is a weighted combination of two distinct distributions. Hence, transitions in bimodal distribution in Fig. 4.9 and 4.10 are visibly distinct from transitions in bimodal distributions for previous fast switching and Schlögl model, where distribution transitions from a unimodal to a bimodal and again to a unimodal regime.

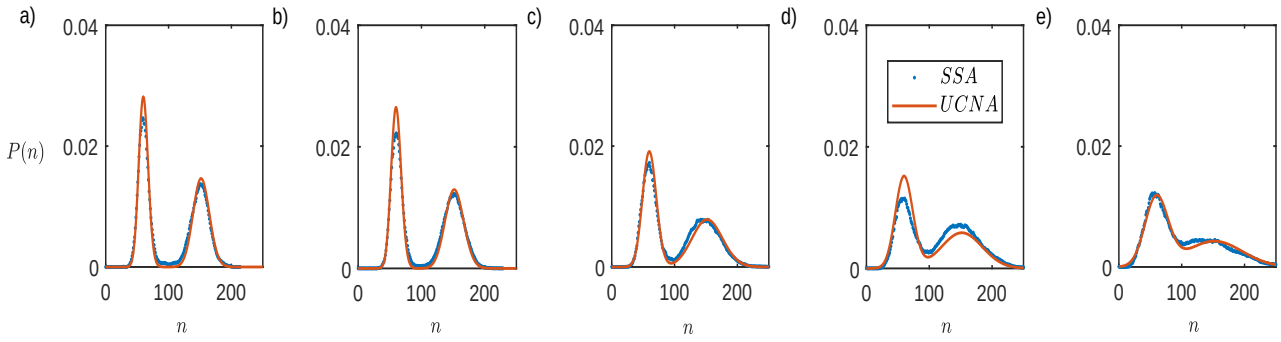

Figure 4.9: For protein translation through a positive feedback and 2-state slow random switching promoter, distribution width around both modes increases with the increase in upstream transcriptional fluctuations and eventually modes merge resulting in loss of bimodality. Stationary state distribution estimated as weighted combination of two UCNA distributions is in good agreement with one extracted using SSA. D = (a) 0.005 (b) 0.01, (c) 0.05, (d) 0.1 and (e) 0.2. Other parameters are $r_2 = 0.5$, $k_0 = 0.15\Omega$, $k_1 = 0.38\Omega$, $p_1 = 0.1$, $p_2 = 0.01$, $\tau = 0.5$ and $\Omega = 200$.

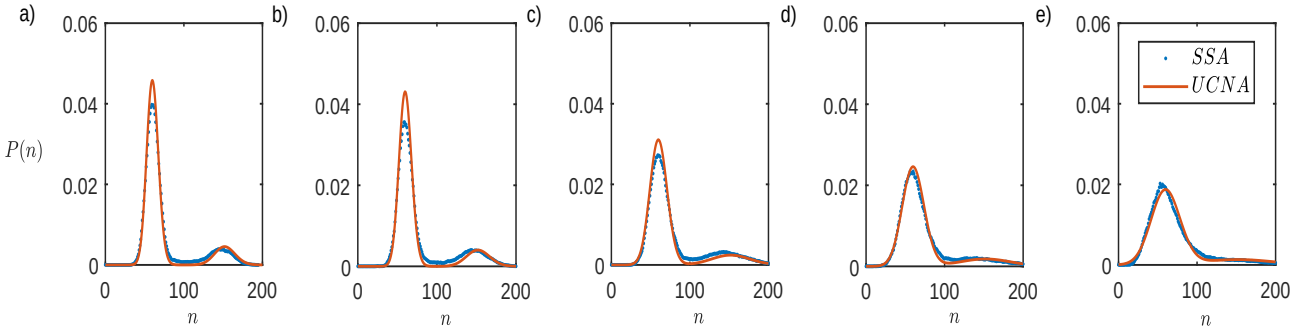

Figure 4.10: For protein translation through a negative feedback and 2-state slow random switching promoter, distribution width around both modes increases with the increase in upstream transcriptional fluctuations and eventually modes merge resulting in loss of bimodality. Stationary state distribution estimated as weighted combination of two UCNA distributions is in good agreement with one extracted using SSA. D = (a) 0.005 (b) 0.01, (c) 0.05, (d) 0.1 and (e) 0.2. Other parameters are $r_2 = 0.5$, $k_0 = 0.38\Omega$, $k_1 = 0.15\Omega$, $p_1 = 0.1$, $p_2 = 0.01$, $\tau = 0.5$ and $\Omega = 200$.

Similar analysis is presented for downstream proteolytic enzyme fluctuations with birth and death rates k_3 and r_3 respectively, where gene expression mechanism can be described as following,

Expression from basal state:

$$Gene \xrightarrow{k_0} Protein + Gene$$
$$Protein \xrightarrow{\frac{1}{E_s} r_2 E} \Phi$$

Expression from activated state:

$$Gene^* \xrightarrow{k_1} Protein + Gene^*$$
$$Protein \xrightarrow{\frac{1}{E_s} r_2 E} \Phi$$

Gene switching:

$$Gene \underset{p_2}{\overset{p_1[Protein]^2}{\rightleftharpoons}} Gene^* \tag{4.26}$$

Analytical approximation based on UCNA and weighted combination of distributions can be evoked with following changes in UCNA terms: For protein n stationary state distribution $P_G(n)$ from gene basal state G:

$$\begin{aligned} h(n) &= k_0 - r_2 n, \\ g_1(n) &= -r_2 n, \\ g_2(n) &= \sqrt{k_0 + r_2 n} \end{aligned} \tag{4.27}$$

$$\text{with} \quad D = \frac{1}{k_3 \Omega}, \quad \tau = \frac{1}{r_3} \text{ and } \lambda = 0$$

and for protein n stationary state distribution $P_{G^*}(n)$ from gene activated state G^*:

$$\begin{aligned} h(n) &= k_1 \Omega - r_2 n, \\ g_1(n) &= -r_2 n, \\ g_2(n) &= \sqrt{k_1 \Omega + r_2 n} \end{aligned} \tag{4.28}$$

$$\text{with} \quad D = \frac{1}{k_3 \Omega}, \quad \tau = \frac{1}{r_3} \text{ and } \lambda = 0$$

Results in the case of slow switching promoter for downstream enzymatic fluctuations as well as for upstream transcription fluctuations, as shown in Fig. 4.11 and 4.12, show distinct characteristics in comparison to analysis presented for fast switching promoter. In this case, low frequency of switching between basal and activated states leads to two distinct distributions, which result in bimodality even in negative feedback conditions.

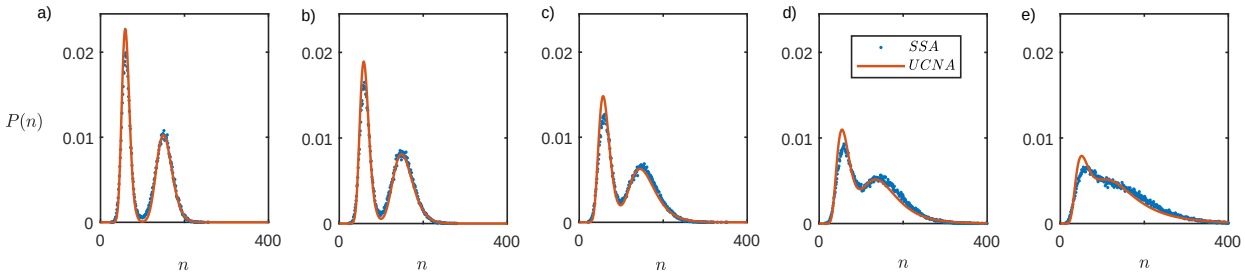

Figure 4.11: For protein translation through a positive feedback and 2-state slow random switching promoter, distribution width around both modes increases with the increase in downstream fluctuations in proteolytic enzyme and eventually modes merge resulting in loss of bimodality. Stationary state distribution estimated as weighted combination of two UCNA distributions is in good agreement with one extracted using SSA. $D =$ (a) 0.025 (b) 0.05, (c) 0.1, (d) 0.2 and (e) 0.375. Other parameters are $r_2 = 0.5$, $k_0 = 0.15\Omega$, $k_1 = 0.38\Omega$, $p_1 = 0.1$, $p_2 = 0.01$, $\tau = 0.5$ and $\Omega = 200$.

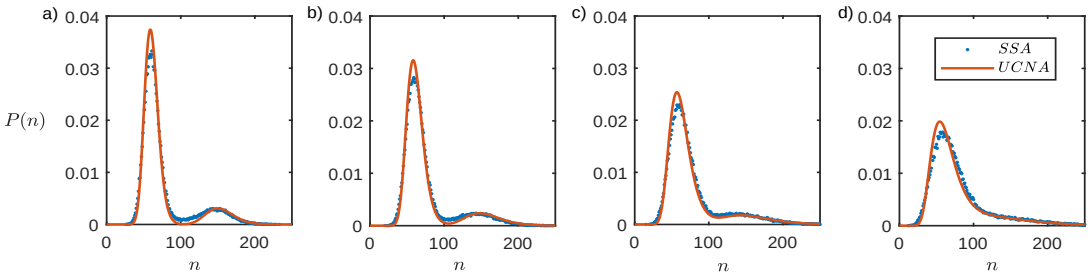

Figure 4.12: For protein translation through a negative feedback and 2-state slow random switching promoter, distribution width around both modes increases with the increase in downstream fluctuations in proteolytic enzyme and eventually modes merge resulting in loss of bimodality. Stationary state distribution estimated as weighted combination of two UCNA distributions is in good agreement with one extracted using SSA. $D =$ (a) 0.025 (b) 0.05, (c) 0.1, (d) 0.2 and (e) 0.375. Other parameters are $r_2 = 0.5$, $k_0 = 0.38\Omega$, $k_1 = 0.15\Omega$, $p_1 = 0.1$, $p_2 = 0.01$, $\tau = 0.5$ and $\Omega = 200$.

4.6 Conclusion

In this chapter, we studied the impact of two distinct external fluctuation sources in gene expression mechanism which have been ignored in most of the exact solutions, derived previously. Inclusion of these new foreign components into models, not only extends predictive strength of models as an inference tool, but also provides new possibilities for design and control in synthetic gene circuits. Along with our analysis on the application of UCNA with Schlögl model, here its applications on various gene expression mechanisms, are adequate examples of how existing methods can be

tinkered and developed as novel tools in emerging dimensions of systems and computational biology.

Chapter 5

CORRELATION RESONANCE UNDER EXTERNAL FLUCTUATIONS

5.1 Introduction

As shown in chapters 3 and 4, external fluctuations can either shift the phenotypic make up of an isogenic population or bring in temporal switching between multiple phenotypes. Similarly, we showed while modelling mesoscopic dynamics of gene expression, incorporating foreign sources of noise, is an effective way to include interactions with intermediate transient catalytic and template molecules such as mRNA. Hence, incorporating external noise provides an opportunity to further extend the reach and prediction strength of these models. In the previous chapters, we studied a few of these examples and saw how external fluctuations can change mesoscopic distributions. These models were either one-variable models or could be reduced to one-variable models under legitimate approximations, as shown with approximation of an exponentially correlated Poisson noise by a Gaussian coloured noise, which thereby reducing a model from two-variables to one-variable. This model reduction approach also made it feasible to extend UCNA to provide approximate

analytical solutions, as shown in chapter 3 and 4 with the use of Schlögl model and different gene expression mechanisms. In this chapter we further extend our analysis on effects of external fluctuations due to interactions with dynamic intermediate molecules for a specific multi-variable model of translation networks.

5.2 Correlation Resonance due to Proteolytic Crosstalk in Translation Networks

An interesting phenomena that emerges from stochastic analysis of one such multi-variable model of gene expression, and is also demonstrated in vivo, is termed as Correlation Resonance. As cells have only limited resources, enzymatic or intermediary components often present bottlenecks in reaction fluxes, which are sometimes ignored in reaction schemes [61], [62]. Correlation Resonance is a peculiar behaviour of mesoscopic correlations between the expression of multiple proteins, due to sharing of a limiting resource in the form of a protein degradation enzyme [61]. Expression of multiple proteins while sharing a single proteolytic degradation enzyme in a translation network (as shown in Fig. 5.1), leads to a correlation between proteins that peaks when the enzyme is utilized to its maximum processing capacity. Multi-class queuing theory provides a theoretical framework where the availability of enzymatic resources considered as 'servers' can lead to correlations between proteins waiting in queues as 'customers' to be served. Depending on translation rates of proteins, enzyme utilization remains either under/over-loaded or reaches a balance point, where the rate of influx of proteins equals the processing capacity of the enzyme and correlations reach a maximum. This analysis paved the way for further experimental design

and verification. Using E.*coli*'s proteolytic enzyme ClpXp as processing *servers* and different labelled fluorescent proteins, this upstream translational crosstalk has been demonstrated in microbial systems [62]. Similarly, an opposite destructive Correlation Resonance due to transcription bottlenecks led by ribosome limitations has been shown in silico [93].

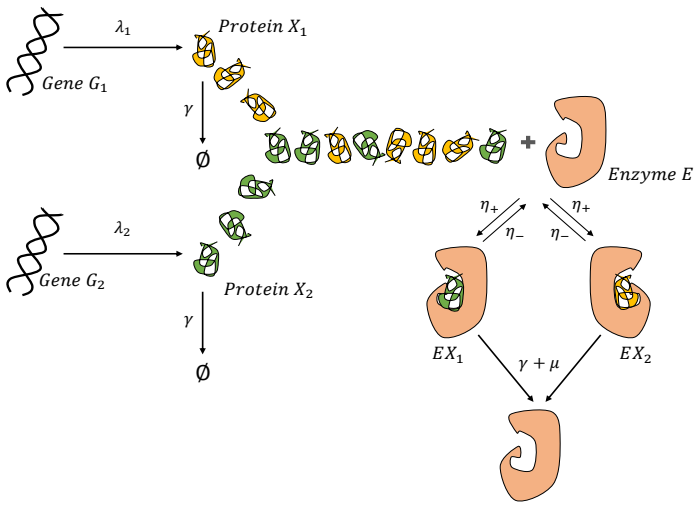

Figure 5.1: Schematic of a translation network which expresses two distinct proteins (P_1, P_2) that share a downstream proteolytic enzyme E. Translation is modelled as a simple birth mechanism without incorporating any upstream transcription.

Though Correlation Resonance emerges solely due to inherent stochasticity of queuing systems, we found it interesting to test by further including simple birth-death transcription mechanisms. Hence, a dynamic upstream source of fluctuations along with downstream bottleneck due to limiting proteolytic enzyme molecules provides further extension on Correlation Resonance. In our analysis, we consider a translation network of two proteins and a shared proteolytic enzyme with restricted availability [62]. External fluctuations are incorporated similarly to chapter 3, as discrete random signals using one or multiple dynamic extrinsic species, affecting translation propensities of one or both proteins while leaving deterministic stationary states unperturbed. We implement the SSA [5] for numerically exact and the LNA [27, 94] for analytically approximate estimation of the

correlation coefficient (CC) between proteins. In results, we demonstrate a new and different downstream Correlation Resonance which can be both constructive or destructive depending upon how the system parameters are tuned.

5.2.1 Correlation Resonance and a Linear Noise Approximation (LNA)

Proteolytic crosstalk in translation was shown as downstream coupling between two proteins through competition for proteolytic enzyme. Eq. 5.1 depicts reaction schema corresponding to translation network in Fig. 5.1. Here, X_1 and X_2 are proteins, with translation rates λ_1 and λ_2. E is proteolytic enzyme, degrading both proteins with same kinetic parameters. γ is dilution due to cell growth and μ is degradation rate due to proteolysis. For each protein, production, degradation and dilution are independent of each other. Similarly, production of each protein is independent of each other. Only shared resource between two proteins is downstream proteolytic enzyme E. Whenever a free enzyme molecule is available, it randomly selects a protein molecule for degradation. We have considered Michaelis-Menten like kinetics for enzymatic degradation of proteins. This leads to upstream correlations relying on availability of enzyme molecules for protein degradation, see Fig. 5.2. Due to increasing mathematical complexity in following cases, instead of using queuing theory model, we adhere with LNA for approximate analytical estimation of CC.

$$\begin{aligned} \phi &\xrightarrow{\lambda_1} X_1 \\ X_1 &\xrightarrow{\gamma} \phi \\ \phi &\xrightarrow{\lambda_2} X_2 \end{aligned}$$

$$
\begin{aligned}
X_2 &\xrightarrow{\gamma} \phi \\
X_1 + E &\xrightarrow{\eta_+} X_1E \\
X_1E &\xrightarrow{\eta_-} X_1 + E \\
X_2 + E &\xrightarrow{\eta_+} X_2E \\
X_2E &\xrightarrow{\eta_-} X_2 + E \\
X_1E &\xrightarrow{\gamma+\mu} E \\
X_2E &\xrightarrow{\gamma+\mu} E
\end{aligned} \tag{5.1}
$$

Fig. 5.2 shows a good agreement between LNA and SSA even for single enzyme molecule kinetics (number of enzyme molecules, $N = 1$) and for multiple enzyme protein binding rates η_+, thus substantiating the use of LNA for a wide range of Michaelis constant $K = \frac{\eta_-}{\eta_+}$, which indicates different regimes of enzymatic affinity. Here, CC is stationary correlation coefficient between protein X_1 and X_2,

$$CC = \frac{\sigma_{X_1,X_2}}{\sigma_{X_1}\sigma_{X_2}}, \tag{5.2}$$

here, σ_{X_1}, σ_{X_2} are stationary state standard deviation of X_1 and X_2 and σ_{X_1,X_2} is stationary state covariance between X_1 and X_2. LNA results are calculated using stationary state solution of Lyapunov Equation shown in chapter 2.

For $\gamma \ll \mu$, correlations between both proteins reach a maximum close to *Balance Point*. Balance point is defined as the point where sum of translation rates of both proteins $\Omega(\lambda_1 + \lambda_2)$ equals available degradation capacity of enzyme $N\mu$, where Ω represents system size and N is number of total enzyme

molecules,

$$\Omega(\lambda_1 + \lambda_2) = N\mu, \tag{5.3}$$

indicating a balance between substrate production rate and maximum processing capacity of the shared enzyme. Similar to [61], we observed a increased correlation peak with decrease in K or increase in η_+ for constant η_-. This suggests correlation maxima is dependent upon affinity of proteolytic enzyme toward proteins. A detailed description on how ratio of two translation rates $\frac{\lambda_1}{\lambda_2}$ further affects peak correlation can also be derived, as shown in [61]. It is important to note that agreement between LNA and SSA as shown in Fig. 5.2 is solely due to symmetry in network topology and because of reported variable being CC. Symmetry arises due to both proteins following same indistinguishable interactions. In contrast, Fig. 5.3 and 5.4 depict deviation between LNA and SSA for model parameters used in Fig. 5.2. Fig. 5.4 shows typical LNA characteristic in form of deviations from SSA due to higher propensity of bimolecular reaction ($\eta_+ = 200$).

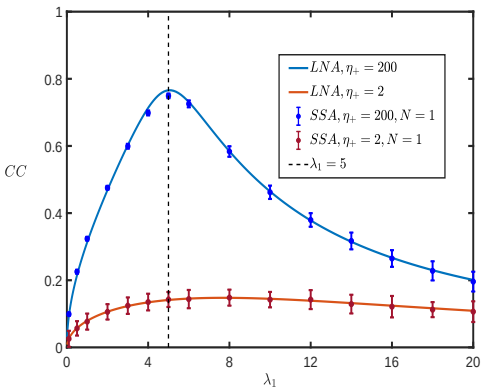

Figure 5.2: Correlation Resonance due to downstream proteolytic crosstalk in translation network from Fig. 5.1 (Eq. 5.1), for different η_+ values. Each dot with error bars corresponds to mean and standard deviation of CC from 100 parallel SSA simulation trajectories. Solid lines represent analytical approximation for CC derived using LNA. $\lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_- = 1000, E_T = 1 (N = 1, \Omega = 1)$

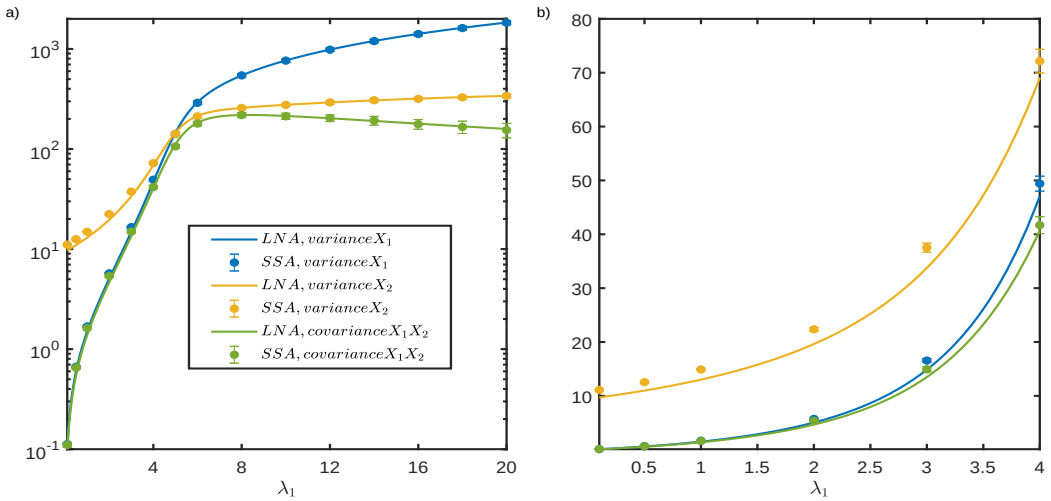

Figure 5.3: Variance and covariance between proteins X_1, X_2 due to proteolytic crosstalk from Fig. 5.1, for $\eta_+ = 2$. Although CC derived using SSA matches with corresponding LNA derivation, this agreement do not follow in case of variance and covariance. $\lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_- = 1000, E_T = 1$. (a) and (b) depict the same data. (b) is the enlarged preview for λ_1 up to 4.

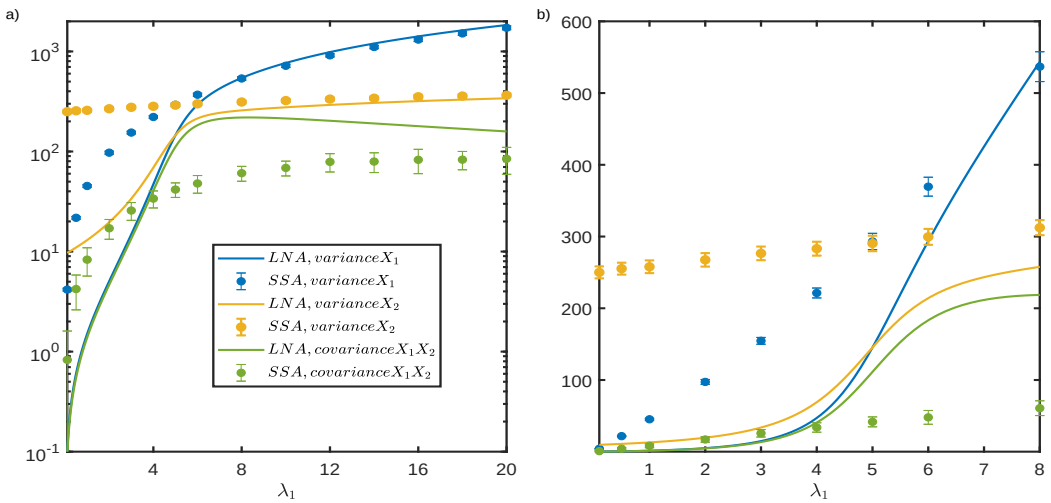

Figure 5.4: Variance and covariance between proteins X_1, X_2 due to proteolytic crosstalk from Fig. 5.1, for $\eta_+ = 200$. Although CC derived using SSA matches with corresponding LNA derivation, this agreement do not follow in case of variance and covariance. $\lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_- = 1000, E_T = 1$. (a) and (b) depict the same data. (b) is the enlarged preview for λ_1 up to 8.

5.3 Correlation Resonance under External Fluctuations

To further examine effects of transcription on protein correlations, we incorporated discrete external fluctuations in translation propensities. We began with incorporating an extrinsic species Y as source of transcription fluctuations on translation of one of the two proteins, X_1. Later on, we compare cases where two independent extrinsic sources influence translation of both proteins independently and translation of both proteins is affected by single extrinsic source.

5.3.1 Case I: Only Y acting on X_1

A species Y is introduced as source of environmental fluctuations on translation of protein X_1, keeping rest of the reactions intact from Eq. 5.1.

$$\begin{aligned} \phi &\underset{r}{\overset{k}{\rightleftharpoons}} Y \\ Y &\xrightarrow{\frac{1}{y_s}\lambda_1} X_1 + Y \end{aligned} \tag{5.4}$$

As shown in Fig. 5.5 and Eq. 5.4, Y follows a simple birth and death kinetics. Kinetics of Y can be inferred as widely popular simple transcription mechanism which has a Poisson distribution with equal mean and variances $(\frac{k\Omega}{r})$ and exponential autocorrelation function with time scale $\tau = \frac{1}{r}$. It is important to note here that, we have considered transcription as a zero order reaction, instead of a first order reaction (as done in chapter 3, Poisson distribution with equal mean and variance $(\frac{k}{r}\frac{1}{\Omega})$). Similarly, to keep deterministic stationary states similar to original, macroscopic translation

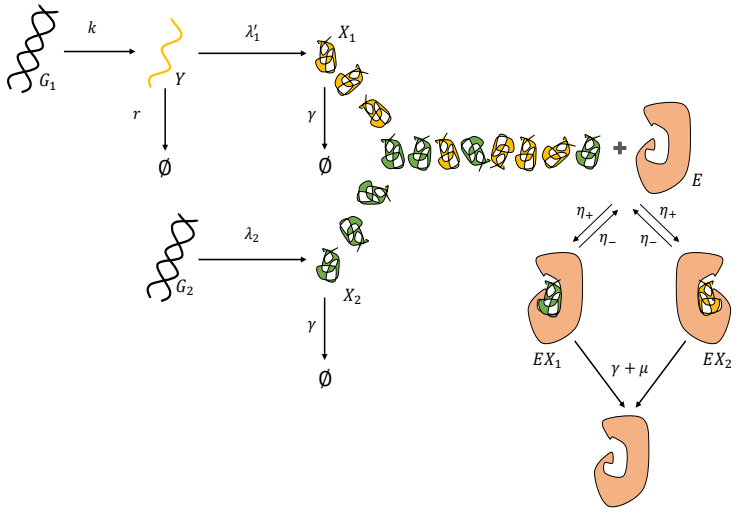

Figure 5.5: In contrast to original schematic shown in Fig. 5.1, we incorporate an upstream transcriptional fluctuation Y which influences the translation of protein X_1, while translation of protein X_2 and downstream proteolytic enzyme sharing remain unchanged.

rate constant λ_1 is divided by stationary state concentration of Y, $y_s = \frac{k}{r}$, leading to a transformed macroscopic translation rate constant λ_1'.

Fig. 5.6 shows SSA and LNA variation in CC, with changing ${k_1}^{-1}$ for $\eta_+ = 2, \lambda_1 = 5$, as CC for original model from Eq. 5.1 is quite small and amplification due to external fluctuations is much more evident. For fixed Ω and r, change in k^{-1} refers to change in fluctuation intensity of Y. While with increase in noise intensity, LNA predicts a sigmoid rise in CC, to our surprise, we observed a distinguishing maximum, which is apparent in reaction systems with low enzyme molecules and faster time scales of external fluctuations. It is evident that presence of upstream external fluctuations on protein X_1, along with downstream proteolytic crosstalk magnifies Correlation Resonance itself. Though LNA fails to predict amplified resonance due to external fluctuations for regimes where mean molecule number of Y ($= \frac{k\Omega}{r}$) is small, yet it provides sufficiently accurate prediction in the range where mean molecule number of Y ($\langle Y \rangle_s = \frac{k\Omega}{r}$) is sufficiently large. Thus, it proves its applicability as an easier mesoscopic

analysis tool for biological systems where finite system size is not a limiting parameter.

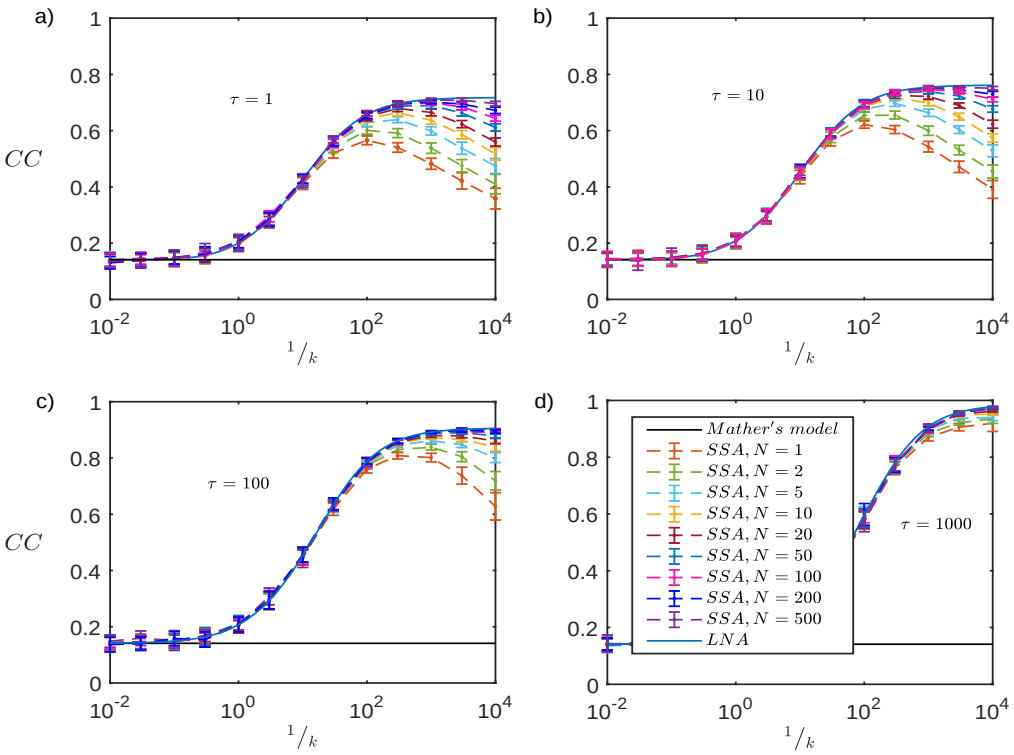

Figure 5.6: Upstream transcription fluctuations Y on protein X_1 amplifies Correlation Resonance. CC further escalates and displays a higher maximum, evident at faster time scales τ, higher intensities $(^1/_k)$ of Y and small proteolytic enzyme molecule numbers N. Each dot with error bars corresponds to CC mean and standard deviation which are extracted from parallel SSA trajectories. Solid black line corresponds to CC maxima at balance point in original mechanism (without Y, Fig. 5.1, Eq. 5.1). Solid blue line is analytical prediction derived using LNA. $\tau =$ (a) 1, (b) 10, (c) 100 and (d) 1000. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

Simulated scaled trajectories of X_1 and X_2 for two different time scales $\tau = 1$ and 100 are shown in Fig. 5.7, corresponding to increasing, maximum and decreasing CC regimes from Fig. 5.6. A mechanistic explanation for amplification of Correlation Resonance and failure of LNA comes from the dependence of bursts size and frequency of X_1 on kinetics of Y. From mechanism shown in Fig. 5.5, it is evident that when a production burst of X_1 (due to bursty kinetics of Y) happens, it occupies available enzyme molecules, which causes accumulation of X_2 as well. In the beginning, when k^{-1} is small enough (for fixed τ, $\langle Y \rangle_s \gg 0$, as shown in first and second

columns of Fig. 5.7), burst frequency is sufficiently large and with increasing burst size, *CC* increases. But as frequency of production bursts decreases further beyond *CC* maxima ($\langle Y \rangle_s$ becomes small), X_1 starts decaying to zero until next burst occurs and X_2 becomes independent of X_1 in that period (as shown in third column of Fig. 5.7), thus *CC* starts decreasing again beyond *CC* maxima. It also explains as relaxation time of X_1 increases (as τ increases), even when burst frequencies are small, *CC* maxima occurs at higher k^{-1} and discrepancies between SSA and LNA decrease even at smaller enzyme counts.

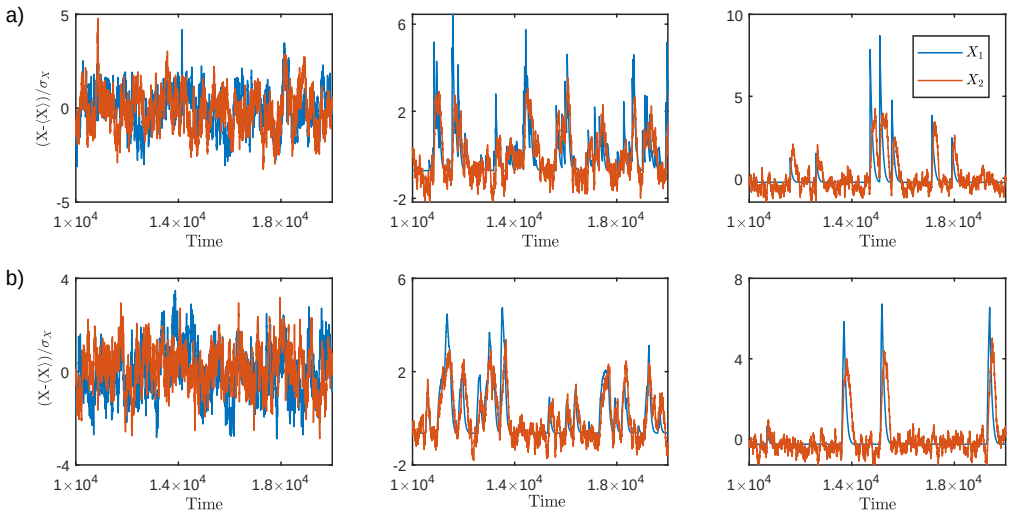

Figure 5.7: SSA trajectories for case I (upstream transcription fluctuations Y on protein X_1) show initial rise and subsequent decrease in *CC* with increase in fluctuation intensity $(1/k)$ for two different time scales, (a) $\tau = 1$ and (b) $\tau = 100$. For (a) $\frac{1}{k} = 0.01, 101.84$ and 1000 and for (b) $\frac{1}{k} = 0.01, 345.62$ and 3000 (left to right). $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1, N = 1$

Fig. 5.8 further explains deviation between *CC* predicted using LNA and the one predicted using SSA. Due to asymmetry of external fluctuation, acting solely on one of the two proteins at smaller system size, dissimilar deviations in LNA variance and SSA variance of X_1 and X_2 provide quantitative explanation of this novel amplified Correlation Resonance.

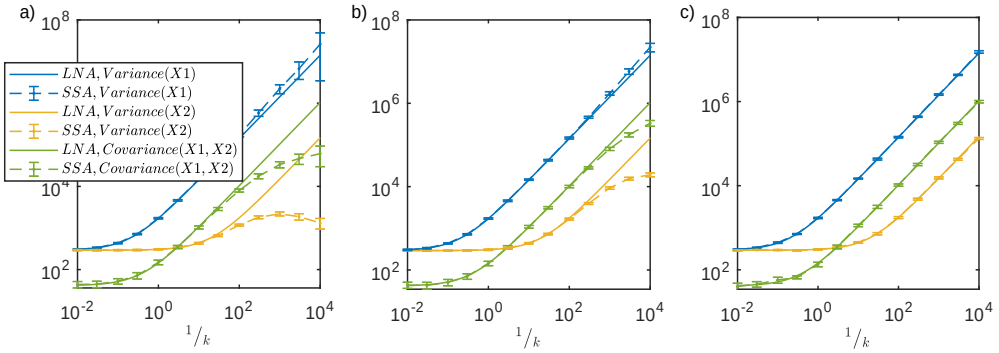

Figure 5.8: Variance and covariance between protein X_1 and X_2 for case I (upstream transcription fluctuations Y on protein X_1) using SSA simulations (dots with error bar) and LNA (solid lines). As expected, with increasing system size Ω, agreement between SSA and LNA increases. In contrast to the original mechanism (Fig. 5.1), due to asymmetry in network topology resulted from the presence of Y only on X_1 (Fig. 5.5), variances of X_1 and X_2 vary differently with increase in noise intensity ($^1/_k$). $\Omega = N =$ (a) 1, (b) 10 and (c) 500. $\tau = 1, \lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

Another interesting aspect that emerges from amplification of Correlation Resonance due to external fluctuations is the power law relationship between position of new amplified maxima and enzyme molecule number (for fixed concentrations) as shown in Fig. 5.9.

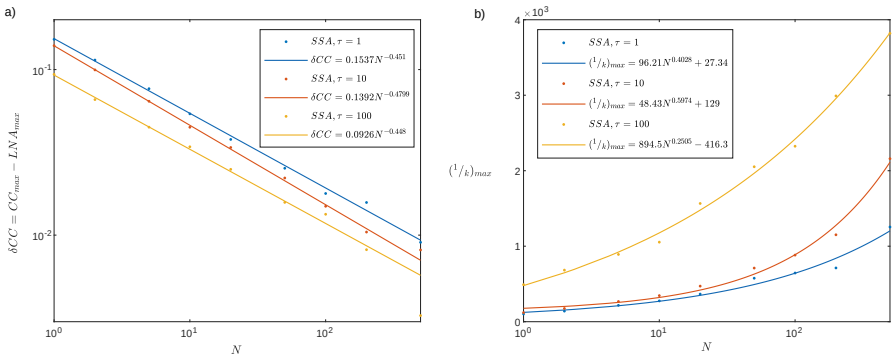

Figure 5.9: A power law relationship is observed between coordinates of CC_{max} (δCC, $^1/_k$) from Fig. 5.6 and number of proteolytic enzyme molecules N for case I (upstream transcription fluctuations Y on protein X_1). Different lines correspond to different time scales of transcription fluctuation τ. Power laws are fitted (a) between δCC (CC_{max} standardized with maximum LNA value) and N (b) between $^1/_k$ corresponding CC_{max} and N. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

It is important to point out here that results in Fig. 5.6 are for constant enzyme concentration $E_t = 1$, hence with changing N, system size Ω also varies accordingly and balance point remains fixed. Hence, in Fig. 5.10 and 5.11, we show similar amplification in CC for fixed system size Ω, referring to a varying balance point with variation in N, for two different time scales τ

and for different values of λ_1 corresponding to different starting CC values, which are smaller, equal and greater than CC at balance point for original system.

Fig. 5.11 indicates similar amplification in correlations due to presence of upstream fluctuations as shown previously in fig. 5.6 for constant enzyme concentration E_T, even during varying translation rate λ_1 where the equality for balance point in equation (5.3) doesn't hold any more.

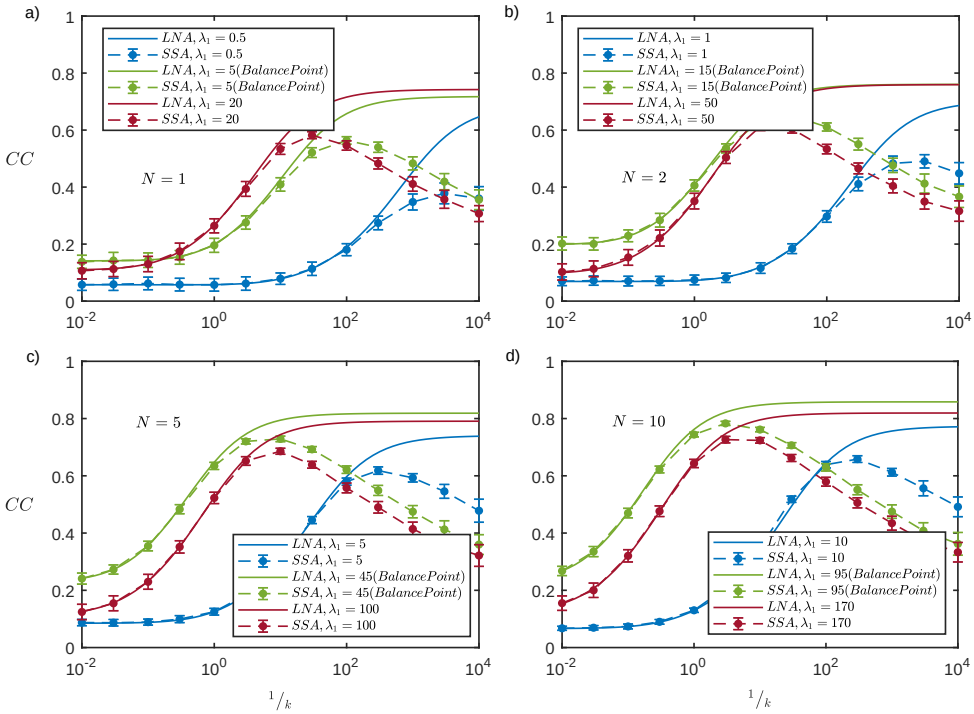

Figure 5.10: Similar amplification of Correlation Resonance as demonstrated in Fig. 5.6, is evident with change in noise intensity ($^1/_k$) and number of proteolytic enzyme molecules N, while system size Ω remains constant instead of constant concentration E_T previously. As balance point (Eq. 5.3) varies with change in N, λ_1 is chosen to depict initial CC (for original mechanism) below, at and above balance point. Dots with standard error are CC mean and standard deviation extracted from parallel SSA trajectories. Solid line are derived using LNA. N = (a) 1, (b) 2, (c) 5 and (d) 10. $\Omega = 1, \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, \tau = 1$.

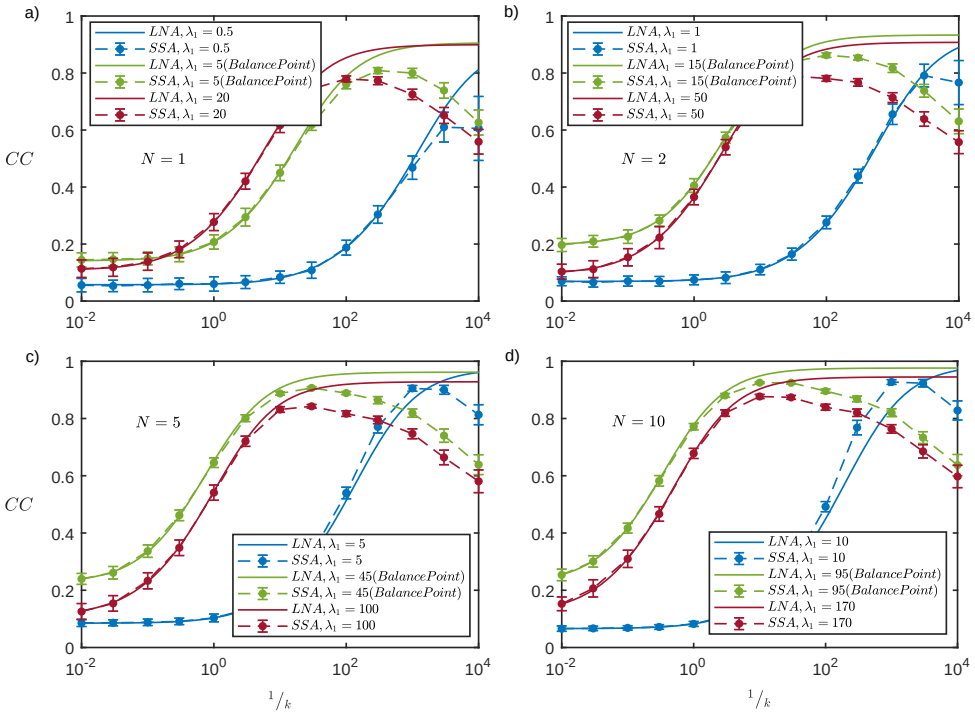

Figure 5.11: Similar amplification of Correlation Resonance as demonstrated in Fig. 5.6, is evident with change in noise intensity ($1/k$) and number of proteolytic enzyme molecules N, while system size Ω remains constant instead of constant concentration E_T previously. As balance point (Eq. 5.3) varies with change in N, λ_1 is chosen to depict initial CC (for original mechanism) below, at and above balance point. Dots with standard error are CC mean and standard deviation extracted from parallel SSA trajectories. Solid line are derived using LNA. $N =$ (a) 1, (b) 2, (c) 5 and (d) 10. $\Omega = 1, \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, \tau = 100$.

5.3.2 Case II: Y and Z acting on X_1 and X_2

Next, along with Y, another species Z following similar kinetics is introduced as source of external fluctuations on both proteins individually as shown in Fig. 5.12 and Eq. 5.5. This modification adds further details in form of transcription noise to both translation propensities λ_1 and λ_2 independently.

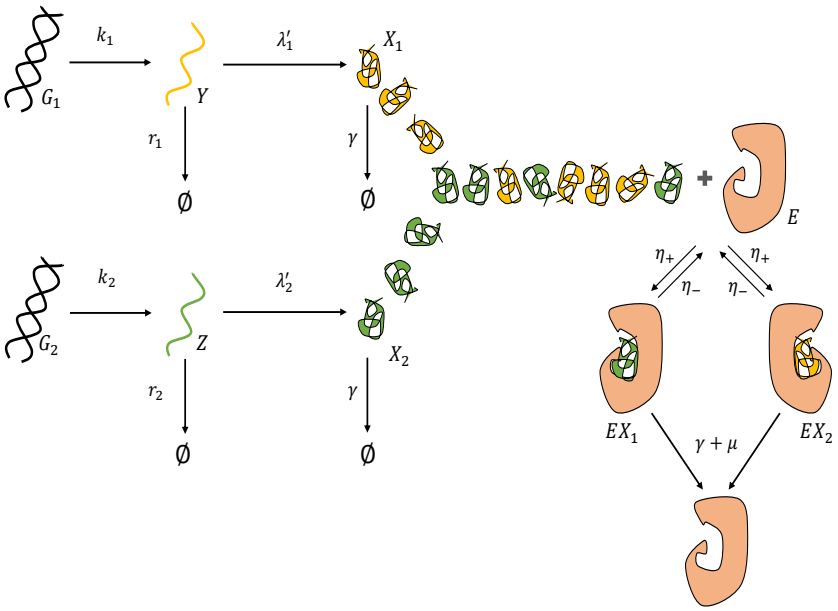

Figure 5.12: A further extension of case I (upstream transcription fluctuations Y on protein X_1) by including two independent upstream transcriptional fluctuation sources Y and Z which influence the translation of protein X_1 and X_2 respectively. Downstream proteolytic enzyme sharing remains unchanged.

$$
\begin{aligned}
\phi &\underset{r_1}{\overset{k_1}{\rightleftharpoons}} Y \\
Y &\xrightarrow{\frac{1}{y_s}\lambda_1} X_1 + Y \\
\phi &\underset{r_1}{\overset{k_2}{\rightleftharpoons}} Z \\
Z &\xrightarrow{\frac{1}{z_s}\lambda_2} X_2 + Z
\end{aligned}
\tag{5.5}
$$

Fig. 5.13 shows LNA predictions for CC between X_1 and X_2 for two different time scales of Y and Z, $\tau_1 = \tau_2 = \tau = 1$ and 1000 respectively. Fig. 5.14 further elaborates both agreement and differences between LNA and SSA for various horizontal slices from Fig. 5.13. As indicated in Fig. 5.14, depending upon how upstream transcription parameters are tuned, we observed a few peculiar changes in correlations between proteins X_1 and X_2 including a minimum where initial correlation is high but with increasing $1/k$, falls back to balance point and rise up again with further increase in $1/k$ (Fig. 5.14, $\tau = 1, k_2 = 10^{-2}, 10^{-1}, N = 500$ and $\tau = 1000, k_2 = 10^{-2}$). These characteristics of correlation curves are driven by change in stochastic kinetics of molecular interactions with changes in upstream transcription parameters and can be explained in the similar way as done in the previous case.

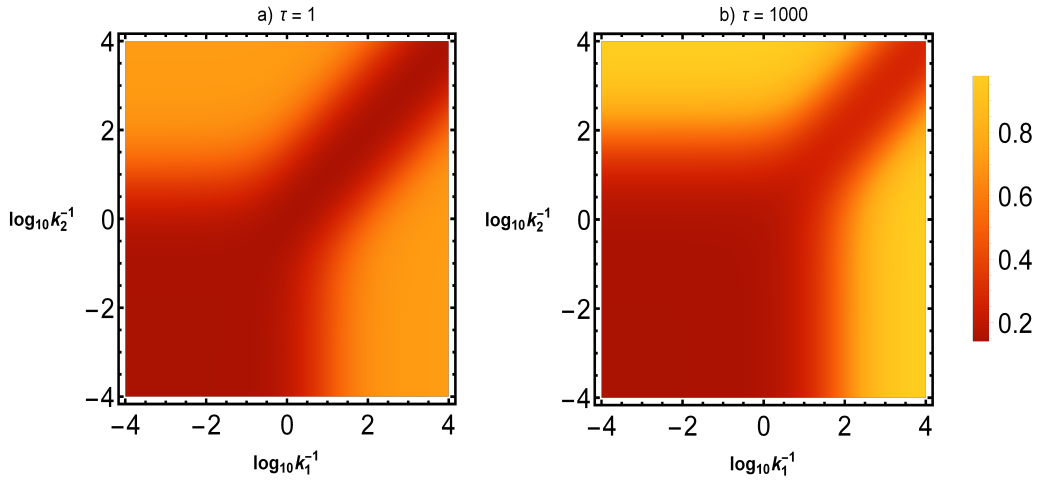

Figure 5.13: LNA derived CC density plots for case II (upstream transcription fluctuations Y and Z acting on translation of X_1 and X_2 respectively). $\tau_1 = \tau_2 = \tau =$ (a) 1 and (b) 1000, where τ_1 and τ_2 are times scales of fluctuations of Y and Z. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

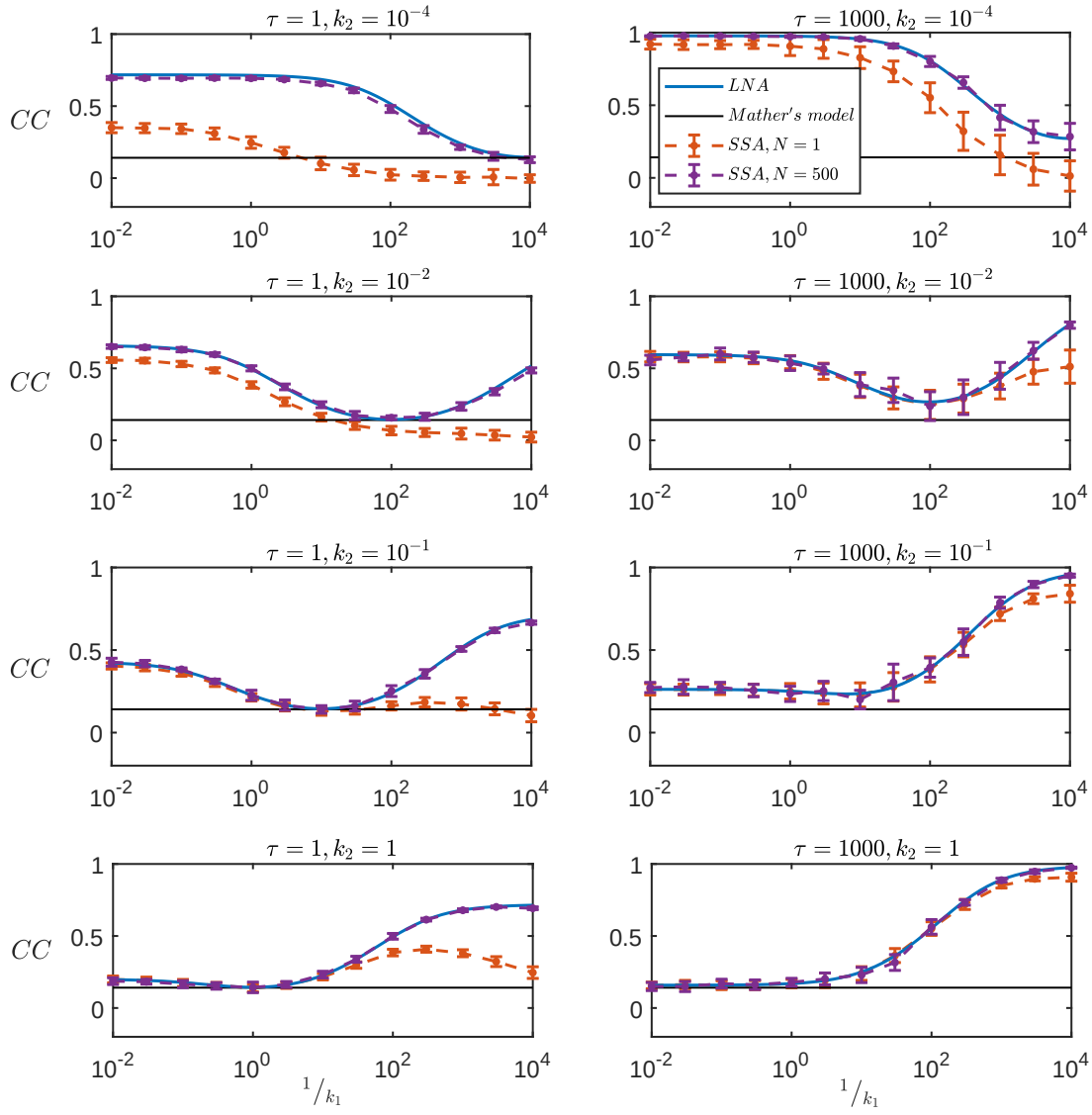

Figure 5.14: Further comparison of horizontal cross-sections of LNA derived CC density map from Fig. 5.13 with SSA simulations for case II (upstream transcription fluctuations Y and Z acting on translation of X_1 and X_2 respectively). It shows different possible manifestations of amplified Correlation Resonance. As expected, agreement between LNA and SSA increases with the increase in system size $\Omega(= N)$, as well as with slower time scale τ. CC is calculated using LNA (solid lines) and SSA (dots with error bars representing CC mean and standard deviation extracted from parallel SSA trajectories), for Y and Z time scales $\tau_1 = \tau_2 = 1$ (column 1) and 1000 (column 2). $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

Quite intuitively, with decreasing fluctuations of one of the two external species, case II tends towards case I, where only one external fluctuation source is considered. This is demonstrated in Fig. 5.15, where different diagonal slices from Fig. 5.13 are plotted along with SSA predictions. Similarity between Fig. 5.15c and Fig. 5.6a, d further establishes this.

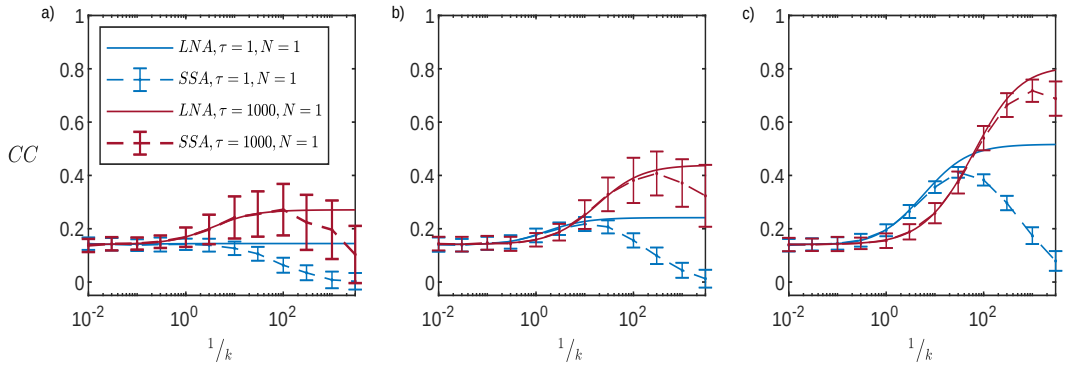

Figure 5.15: Comparison of diagonal cross-sections of LNA derived CC density map from Fig. 5.13 with SSA simulations for case II (upstream transcription fluctuations Y and Z acting on translation of X_1 and X_2 respectively). As one of the two transcription fluctuations weakens, case II exhibits similarities to case I (only Y acting on protein X_1 alone). CC calculated using LNA (solid lines) and SSA (dots with error bars representing CC mean and standard deviation extracted from parallel SSA trajectories), for Y and Z time scales $\tau_1 = \tau_2 = 1$ (blue) and 1000 (red), for different intensity ratios: (a) $\frac{k_1}{k_2} = 1$, (b) $\frac{k_1}{k_2} = 0.1$ and (c) $\frac{k_1}{k_2} = 0.01$. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

5.3.3 Case III: Y acting on both X_1 and X_2

Here, only Y which follows simple birth and death kinetics, is introduced as source of environmental fluctuations on both proteins, as shown in Eq. 5.6 and Fig. 5.16.

$$\begin{aligned} \phi &\underset{r}{\overset{k}{\rightleftharpoons}} Y \\ Y &\xrightarrow{\frac{1}{y_s}\lambda_1} X_1 + Y \\ Y &\xrightarrow{\frac{1}{y_s}\lambda_2} X_2 + Y \end{aligned} \tag{5.6}$$

This scenario can be considered as modification of previous case as Y and Z become completely correlated instead of being independent random species.

Biologically, these mechanisms of extrinsic fluctuations can be mapped to translation from polycistronic mRNAs or widely found operons [95]. Operons provide a single transcript as an expression template for multiple proteins or multiple subunits of a single protein/enzyme.

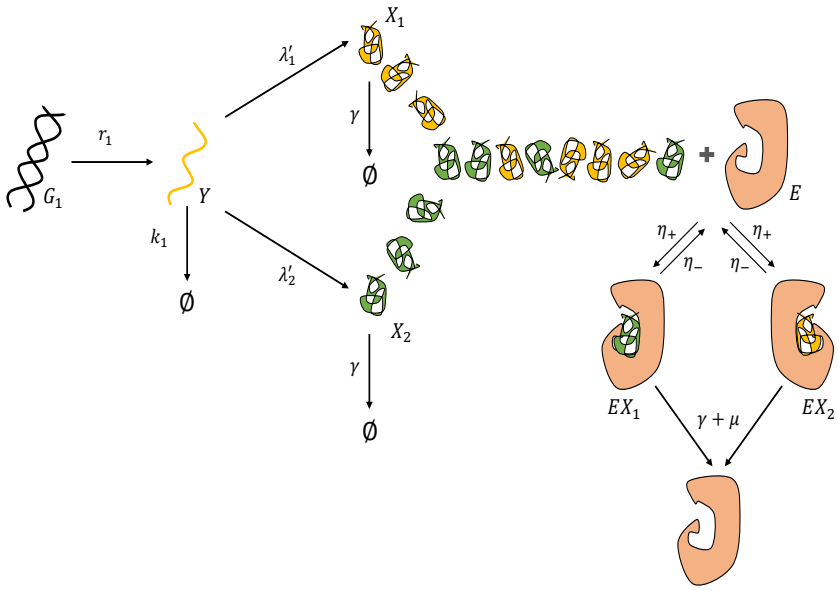

Figure 5.16: A variation of case II where only a single upstream transcriptional fluctuation source Y influences the translation of both protein X_1 and X_2. Downstream proteolytic enzyme sharing remains unchanged.

Surprisingly, LNA results are in good agreement with SSA, even when degradation is led by one single enzyme molecule $N = 1$, see Fig. 5.17. As CC follows a sigmoid curve, it points toward evolution of operons as a regulatory mechanism which provides sufficient synchronization between translation of multiple proteins. In comparison to case I and II, similar to the original model, symmetry in network topology as shown in the form of variances in Fig. 5.19 leads to good agreement between LNA and SSA CC estimations independent of time scale τ or mean $\langle Y \rangle_s$. Fig. 5.18 shows sample trajectories of proteins X_1 and X_2, influenced from a common fluctuation source for multiple time scales.

In case II, we saw how two completely independent external noise sources

changed stochastic dynamics of translation network. In contrast, case III portrays a different picture. Hence, next we present a hypothetical interpolation of two cases, where correlation between Y and Z varies between 0 and 1, mimicking dynamics in between case II and III.

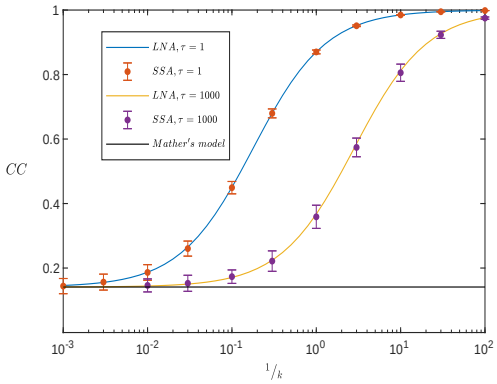

Figure 5.17: In case III (common upstream transcription fluctuation Y influencing translation of both proteins), CC shows a sigmod rise as fluctuation intensity $(^1/_k)$ increases, in contrast to previous two cases where we saw a subsequent decline after an initial peak in CC. LNA (solid colored lines) derived CC curves show surprisingly good agreement with SSA simulations (dots with error bars corresponding to CC mean and standard deviation extracted from parallel SSA trajectories), independent of system size Ω $(= N)$. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

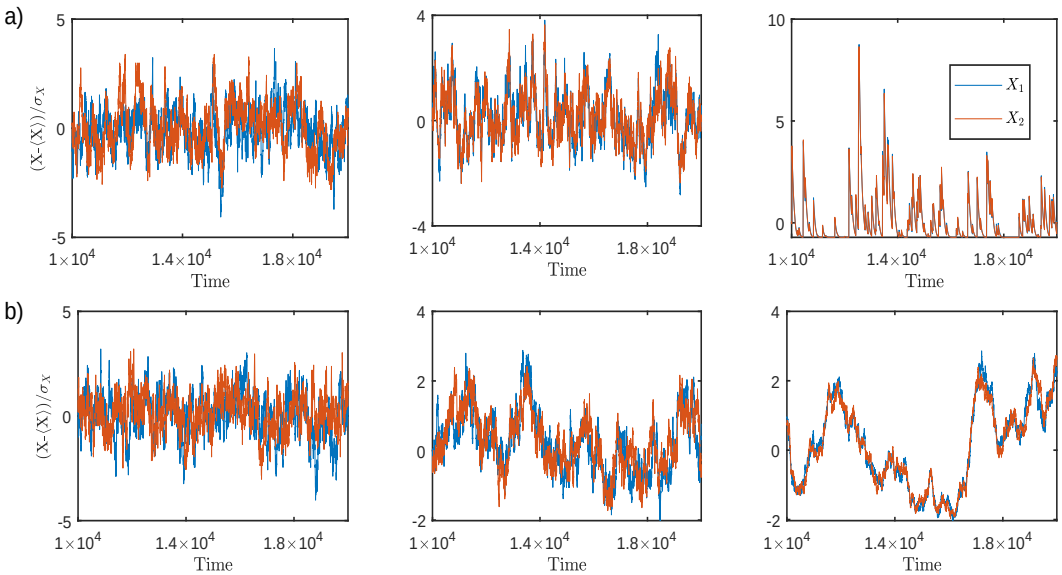

Figure 5.18: SSA trajectories for case III (common upstream transcription fluctuation Y influencing translation of both proteins) confirm the CC between two proteins increases as the intensity of common transcription fluctuations increases. (a) $\tau = 1$ and (b) $\tau = 100$. For (a) $\frac{1}{k} = 0.01, 101.84$ and 1000 and for (b) $\frac{1}{k} = 0.01, 345.62$ and 3000 (left to right). $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1, N = 1$

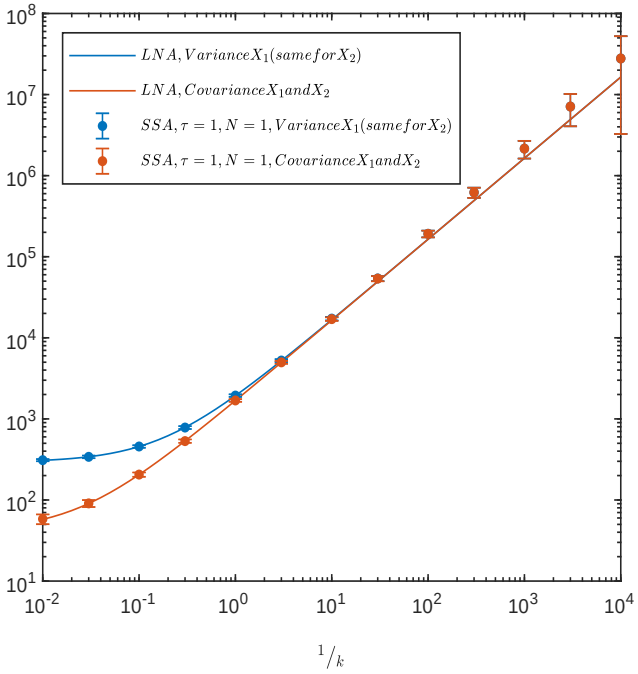

Figure 5.19: Due to symmetry in network topology, both proteins have exactly the same stationary state variance for case III (common upstream transcription fluctuation Y influencing translation of both proteins). Solid lines are plotted from LNA derived analytical expression and dots with error bars are CC mean and standard deviations extracted from parallel SSA trajectories. $\Omega = 1$, $\tau = 1$. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

5.3.4 Hypothetical interpolation between Case II and III

A simple approximate method to introduce correlation between Y and Z is to make artificially correlated Wiener processes which drive LNA derived SDEs, corresponding to FPE. To begin with, for case II, we have external fluctuations in form of Y and Z which follow,

$$d\varepsilon_{n_y} = -r_1 \varepsilon_{n_y} dt + \sqrt{2k_1} dW_t^1, \tag{5.7}$$

$$d\varepsilon_{n_z} = -r_2 \varepsilon_{n_z} dt + \sqrt{2k_2} dW_t^2, \tag{5.8}$$

In context of case III, Y and Z are exactly alike. In general, stationary state covariance between variables ε_{n_y} and ε_{n_z} can be written as,

$$\sigma_{\varepsilon_{n_y},\varepsilon_{n_z}} = \frac{2\sqrt{k_1 k_2}\rho}{r_1 + r_2}, \tag{5.9}$$

where ρ is CC between Wiener processes W_t^1 and W_t^2, and is 0 and 1 for case II and case III, respectively. Hence with varying value of ρ, we can achieve intermediate correlation between Y and Z, which we are interested in. Fig. 5.20 depicts gradual increment CC starting from base value at case II up to highest value at case III, with varying ρ. CC is calculated by simulation of SDEs (as shown in Eq. 5.10) corresponding to case II but with non-zero correlation between Y and Z. Hence, a full set of SDEs capturing dynamics of fluctuations $\boldsymbol{\varepsilon} = [\varepsilon_{n_y}; \varepsilon_{n_z}; \varepsilon_{n_{X_1}}; \varepsilon_{n_{X_2}}; \varepsilon_{n_E}; \varepsilon_{n_{EX_1}}]$ around deterministic stationary state concentrations $([Y]_s, [Z]_s, [X_1]_s, [X_2]_s, [EX_1]_s, [E]_s)$ for participating constituents in mechanism from case II, can be written as,

$$d\boldsymbol{\varepsilon} = \mathbf{J}\boldsymbol{\varepsilon} + \boldsymbol{\sigma} d\mathbf{W}_t, \tag{5.10}$$

where, $\mathbf{W} = [W_t^1; W_t^2; W_t^3; W_t^4; W_t^5; W_t^6]$, represents a vector of Wiener processes. $\mathbf{J}$ is Jacobian matrix and $\boldsymbol{\sigma}$ denotes the diffusion term derived from diffusion matrix $\mathbf{D}$ as $\mathbf{D} = \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}^T$, which are defined as following,

$$\mathbf{J} = \begin{pmatrix} J_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & J_{22} & 0 & 0 & 0 & 0 \\ J_{31} & 0 & J_{33} & 0 & J_{35} & J_{36} \\ 0 & J_{42} & 0 & J_{44} & J_{45} & J_{46} \\ 0 & 0 & J_{53} & J_{54} & J_{55} & 0 \\ 0 & 0 & J_{63} & 0 & J_{65} & J_{66} \end{pmatrix} \tag{5.11}$$

$$\mathbf{D} = \begin{pmatrix} D_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & D_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & D_{33} & 0 & D_{35} & D_{36} \\ 0 & 0 & 0 & D_{44} & D_{45} & 0 \\ 0 & 0 & D_{53} & D_{54} & D_{55} & D_{56} \\ 0 & 0 & D_{63} & 0 & D_{65} & D_{66} \end{pmatrix} \tag{5.12}$$

For Eq. 5.11, we have,

$$\begin{aligned} J_{11} &= -r_1, & J_{22} &= -r_2, \\ J_{31} &= \lambda_1', & J_{33} &= -\gamma - [E]_s\eta_+, \\ J_{35} &= -[X_1]_s\eta, & J_{36} &= \eta_-, \\ J_{42} &= \lambda_2', & J_{44} &= -\gamma - [E]_s\eta_+, \\ J_{45} &= -\eta_- - [X_2]_s\eta_+, & J_{46} &= -\eta_-, \\ J_{53} &= -[E]_s\eta, & J_{54} &= -[E]_s\eta_+, \\ J_{55} &= -\gamma - \mu - \eta_- - [X_1]_s\eta_+ - [X_2]_s\eta_+, & & \\ J_{63} &= [E]_s\eta_+, & J_{65} &= [X_1]_s\eta_+, \\ J_{66} &= -\gamma - \mu - \eta_-, & & \end{aligned}$$

and for Eq. 5.12, we have,

$$\begin{aligned} D_{11} &= k_1 + r_1 y_s, \\ D_{22} &= k_2 + r_2 z_s, \\ D_{33} &= [X_1]_s\gamma + [EX_1]_s\eta_- + [E]_s[X_1]_s\eta_+ + \lambda_1' y_s, \\ D_{35} &= [EX_1]_s\eta_- + [E]_s[X_1]_s\eta_+, \\ D_{36} &= -[EX_1]_s\eta_- - [E]_s[X_1]_s\eta, \\ D_{44} &= [X_2]_s\gamma + (-[EX_1]_s - [E]_s + [E_T])\eta_- + [E]_s[X_2]_s\eta_+ + \lambda_2' z_s, \\ D_{45} &= (-[EX_1]_s - [E]_s + [E_T])\eta_- + [E]_s[X_2]_s\eta_+, \end{aligned}$$

$$
\begin{aligned}
D_{53} &= [EX_1]_s\eta_- + [E]_s[X_1]_s\eta_+, \\
D_{54} &= (-[EX_1]_s - [E]_s + [E_T])\eta_- + [E]_s[X_2]_s\eta_+, \\
D_{55} &= [EX_1]_s(\gamma+\mu) + (-[EX_1]_s - [E]_s + [E_T])(\gamma+\mu) + [EX_1]_s\eta_- \ldots \\
&\ldots + (-[EX_1]_s - [E]_s + [E_T])\eta_- + [E]_s[X_1]_s\eta_+ + [E]_s[X_2]_s\eta_+, \\
D_{56} &= -[EX_1]_s(\gamma+\mu) - [EX_1]_s\eta_- - [E]_s[X_1]_s\eta_+, \\
D_{63} &= -[EX_1]_s\eta_- - [E]_s[X1]_s\eta_+, \\
D_{65} &= -[EX_1]_s(\gamma+\mu) - [EX_1]_s\eta_- - [E]_s[X_1]_s\eta_+, \\
D_{66} &= [EX_1]_s(\gamma+\mu) + [EX_1]_s\eta_- + [E]_s[X_1]_s\eta_+
\end{aligned}
$$

Eq. 5.7 and 5.8 are included as first two rows of Eq. 5.10. LNA derived SDEs for the whole network including external fluctuations with different values of ρ, are simulated using Euler Maruyama method [96]. Mean and standard deviations of stationary state CC are reported using data from 10 parallel trajectories. It is to keep in mind that this analysis is only valid for scenarios where LNA is in good agreement with SSA, corresponding to large Ω.

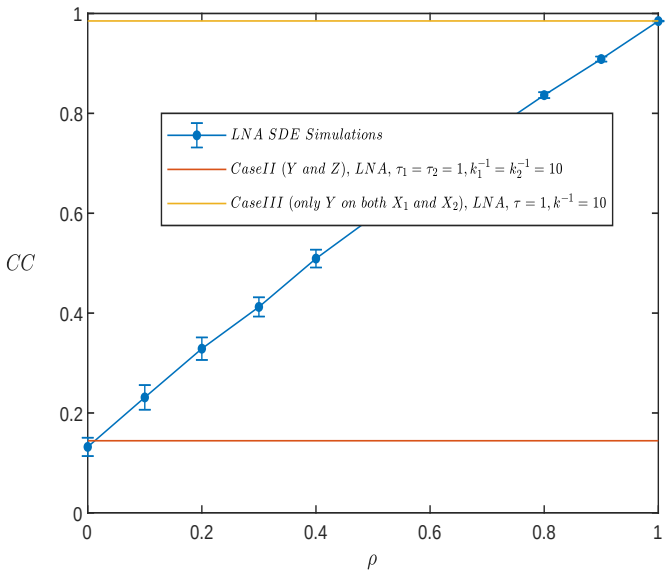

Figure 5.20: A linear relationship which links case II and case III. Dots with error bars denote CC mean and standard deviations extracted from parallel trajectories of LNA derived SDEs, with artificially introduced correlations between Y and Z. $\Omega = 1$, $\tau = 1$. $\lambda_1 = \lambda_2 = 5, \gamma = 0.01, \mu = 10, \eta_+ = 2, \eta_- = 1000, E_T = 1$

Due to fixed amount of cellular resources, a complete statistical independence between different transcripts would be rare to find. Hence, results shown fig. 5.20 are quite intuitive and captures the intermediate correlations in multi-protein translation due to upstream correlations in transcription, where interactions and correlations between different transcripts remain hard to depict through direct interactions.

5.4 Conclusion

Similar to previous chapters, we demonstrated another example of how and why interaction with external fluctuations can drastically change innate mesoscopic dynamics of biological pathways. Analysis of case II provides quantitative foundation for developing tunable synthetic controls of multi-protein translation networks. Stochastic control of correlations in synthetic expression networks by tuning transcription provides alternative routes and a new dimension to make synthetic biology more predictable and robust. Case III caters insights on evolution and necessity of operons. Overall, an important lesson is to consider applicability of these representative schemes for modelling complex biochemical processes with a pinch of salt. There are interactions which are often ignored for mathematical simplifications during modelling and simulation, which can lead to surprising experimental deviations.

Chapter 6

DISCUSSION

In this thesis, we explored how stochastic behaviour (limited to stationary state) of different pathways changes due to interactions with stochastic dynamics of external components. We started with a detailed revision of different simulation and analytical approximation methods used in this work. Next, we explored stochastic dynamics and transitions in stationary state bimodal distributions of Schlögl Model, through different reaction channels. In this analysis, we set up UCNA in combination with model reduction using LNA (OU process) as our primary tool for analytical investigation. Use of UCNA proved helpful in analytical sensitivity analysis and quantification of changes in distribution modes with respect to external noise intensity and time scale over a wide range of system parameters. With UCNA in the picture, we further tested its application as an approximate analytical tool to capture stationary state protein distribution in different gene expression mechanisms. Not only were we able to extend the reach of these gene expression models which generally do not account for mRNA or enzymatic fluctuations, but also due to the simplicity of UCNA, we establish it as a considerably accurate method to provide significant analysis for the design of novel synthetic biology pathways. In chapter 5, we studied effects of transcriptional fluctuations in a translation network with multiple

proteins, which share a downstream proteolytic enzyme. In a way, this study is a specific extension of our analysis in chapter 4, where translation is modelled as a simple birth-death mechanism with an extra downstream temporal constituent. Because the network consists of multiple species, we switch to LNA as our method of choice. We demonstrated various interesting phenomena which a pathway of this complexity can exhibit, such as amplification and destruction of Correlation Resonance with change in noise intensity. We also found evidences of sigmoid rise in correlation between proteins expressed through an operon. Similarly, we explored transcription as a synthetic control in gene expression pathways.

In our analysis, we showed how discrete birth-death mechanisms can be approximated with Gaussian coloured noise. After this transformation, our work finds equivalence with previous studies in this domain, where external fluctuations were imposed on reaction propensities in different forms including Gaussian Normal or log-Normal coloured noise [55, 73, 74, 97] as well as other forms such as Levy processes [98–100]. Initial studies in this context, began with modifying SSA for sampling stochastic trajectories of reaction networks with fluctuating propensities. These studies also discuss reasoning behind choosing external fluctuations on reaction propensities as Gaussian coloured noise. The only factor which limits the choice of random processes as a source of external fluctuations is that reaction propensities should not become negative with different realizations of imposed random processes. A simple trick to avoid this mathematical entanglement is to use external fluctuations which are always positive or have bounded form, unlike Normal distribution. For example, instead of using Normally distributed Gaussian coloured noise, it was recommended to use log-Normal distribution for external fluctuations, especially if significant amount of cumulative probability corresponds to negative values. Because of their

bounded nature, same reasoning applies to use of Levy processes as external fluctuations over reaction propensities. Different variations of SSA have been proposed which successfully account for noise in reaction propensities [55, 101–104]. In contrast, in our work, we deal with a fully discrete system without any initial assumption of inherent distribution about external fluctuations. Instead, we show how a simple birth-death process take a form of Gaussian coloured noise. While implementing UCNA, we approximate external fluctuations with similar Gaussian coloured noise as well, still our analysis avoids any requirements to implement other specific algorithms beyond SSA and as well as any initial assumption about external fluctuations.

Similarly, in the context of systems biology, while main emphasis has always been on deriving closed form solutions without any system size limitations, this approach is always constricted due to the complexity of networks. Hence, our work highlights contributions of CFPE, UCNA and LNA in sequence to further extend the analysis capacity of gene expression models.

Results derived in chapter 3 show the accuracy of UCNA, even for a highly non-linear model. In addition, results from chapter 4, further cement our confidence in applications of UCNA to systematic analysis of gene expression mechanisms. Hence, as a future prospect, it would be ideal to test the robustness of UCNA during parameter inference as well as extend its purview to multivariable models. As an extension to CFPE and CLE, investigations on applications of UCNA in systems biology will be fruitful.

On the other side, since its first application in biological context [27] and its recent extension to reaction systems with fluctuating propensities [105], popularity of LNA as a theoretical analysis tool, has grown exponentially. Another similar work, focusing on effects of mRNA sequestering with the

help of miRNAs and miRNAs as source of external fluctuations in bimodal gene expression also implements LNA as a theoretical tool [106].

With our initial focus on developing novel approximation methods for stochastic analysis of intracellular pathways, we have established a new dimension with simplicity in methods and novelty in analysis. Ultimately, we consider our efforts as incremental steps towards making intracellular biology more predictable and quantifiable.

BIBLIOGRAPHY

[1] Lev S Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2):026601, 2014.

[2] John RS Newman, Sina Ghaemmaghami, Jan Ihmels, David K Breslow, Matthew Noble, Joseph L DeRisi, and Jonathan S Weissman. Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840, 2006.

[3] Yuichi Taniguchi, Paul J Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X Sunney Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, 2010.

[4] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.

[5] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[6] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

[7] Daniel T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.

[8] C Gardiner. Stochastic methods: a handbook for the natural and social sciences 4th ed.(2009).

[9] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of chemical physics*, 124(4):044104, 2006.

[10] Chetan Gadgil, Chang Hyeong Lee, and Hans G Othmer. A stochastic analysis of first-order reaction networks. *Bulletin of mathematical biology*, 67(5):901–946, 2005.

[11] Stephen Smith and Vahid Shahrezaei. General transient solution of the one-step master equation in one dimension. *Physical Review E*, 91(6):062119, 2015.

[12] Ramon Grima. Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems. *Physical Review E*, 92(4):042124, 2015.

[13] Ivan G Darvey, BW Ninham, and PJ Staff. Stochastic models for second-order chemical reaction kinetics. the equilibrium state. *The journal of chemical physics*, 45(6):2145–2155, 1966.

[14] Vahid Shahrezaei and Peter S Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, 105(45):17256–17261, 2008.

[15] Ramon Grima, Deena R Schmidt, and Timothy J Newman. Steady-state fluctuations of a genetic feedback loop: An exact solution. *The Journal of chemical physics*, 137(3):035104, 2012.

[16] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[17] JE Moyal. Stochastic processes and statistical physics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):150–210, 1949.

[18] Daniel T Gillespie. The chemical langevin and fokker- planck equations for the reversible isomerization reaction. *The Journal of Physical Chemistry A*, 106(20):5063–5071, 2002.

[19] Daniel T Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[20] Raya Khanin and Desmond J Higham. Chemical master equation and langevin regimes for a gene transcription model. *Theoretical Computer Science*, 408(1):31–40, 2008.

[21] Desmond J Higham. Stochastic ordinary differential equations in applied and computational mathematics. *IMA journal of applied mathematics*, 76(3):449–474, 2011.

[22] Linda JS Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, 2017.

[23] NG van Kampen. A power series expansion of the master equation. *Canadian Journal of Physics*, 39(4):551–567, 1961.

[24] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[25] Edward WJ Wallace. A simplified derivation of the linear noise approximation. *arXiv preprint arXiv:1004.4280*, 2010.

[26] E Wallace, DT Gillespie, KR Sanft, and LR Petzold. A new perspective on the linear noise approximation. *Preprint (submitted)*, 2013.

[27] Johan Elf and Måns Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome research*, 13(11):2475–2484, 2003.

[28] Matthew Scott, Brian Ingalls, and Mads Kaern. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(2):026107, 2006.

[29] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Stochastic analysis of chemical reaction networks using linear noise approximation. *Biosystems*, 149:26–33, 2016.

[30] Michał Komorowski, Bärbel Finkenstädt, Claire V Harper, and David A Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC bioinformatics*, 10(1):343, 2009.

[31] Paul Fearnhead, Vasilieos Giagos, and Chris Sherlock. Inference for reaction networks using the linear noise approximation. *Biometrics*, 70(2):457–466, 2014.

[32] Stefan Engblom. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation*, 180(2):498–515, 2006.

[33] Colin S Gillespie. Moment-closure approximations for mass-action models. *IET systems biology*, 3(1):52–58, 2009.

[34] Patrick Smadbeck and Yiannis N Kaznessis. A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences*, 110(35):14261–14265, 2013.

[35] Ramon Grima. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *The Journal of chemical physics*, 136(15):04B616, 2012.

[36] David Schnoerr, Guido Sanguinetti, and Ramon Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 143(18):11B610_1, 2015.

[37] Philipp Thomas, Arthur V Straube, and Ramon Grima. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC systems biology*, 6(1):39, 2012.

[38] Stephen Smith, Claudia Cianci, and Ramon Grima. Model reduction for stochastic chemical systems with abundant species. *The Journal of chemical physics*, 143(21):12B615_1, 2015.

[39] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[40] Yang Cao, Daniel T Gillespie, and Linda R Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of chemical physics*, 124(4):044109, 2006.

[41] Darren J Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2011.

[42] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.

[43] Adam Arkin, John Ross, and Harley H McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

[44] William J Blake, Mads Kærn, Charles R Cantor, and James J Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633, 2003.

[45] Mukund Thattai and Alexander Van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530, 2004.

[46] Leor S Weinberger, John C Burnett, Jared E Toettcher, Adam P Arkin, and David V Schaffer. Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuations drive phenotypic diversity. *Cell*, 122(2):169–182, 2005.

[47] Mary J Dunlop, Robert Sidney Cox III, Joseph H Levine, Richard M Murray, and Michael B Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature genetics*, 40(12):1493, 2008.

[48] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[49] Paul C Bressloff. Stochastic switching in biology: from genotype to phenotype. *Journal of Physics A: Mathematical and Theoretical*, 50(13):133001, 2017.

[50] Thomas J Snowden, Piet H van der Graaf, and Marcus J Tindall. Methods of model reduction for large-scale biological systems: a survey of current methods and trends. *Bulletin of mathematical biology*, 79(7):1449–1486, 2017.

[51] Jeremy Gunawardena. Time-scale separation–michaelis and menten's old idea, still bearing fruit. *The FEBS journal*, 281(2):473–488, 2014.

[52] Jae Kyoung Kim, Krešimir Josić, and Matthew R Bennett. The relationship between stochastic and deterministic quasi-steady state approximations. *BMC systems biology*, 9(1):87, 2015.

[53] Don W Cleveland and Tim J Yen. Multiple determinants of eukaryotic. *New Biologist*, 1(2):121–126, 1989.

[54] Philip Mitchell and David Tollervey. mrna stability in eukaryotes. *Current opinion in genetics & development*, 10(2):193–198, 2000.

[55] Vahid Shahrezaei, Julien F Ollivier, and Peter S Swain. Colored extrinsic fluctuations and stochastic gene expression. *Molecular systems biology*, 4(1), 2008.

[56] Christoph Zechner and Heinz Koeppl. Uncoupled analysis of stochastic reaction networks in fluctuating environments. *PLoS computational biology*, 10(12):e1003942, 2014.

[57] Peter Jung and Peter Hänggi. Dynamical systems: a unified colored-noise approximation. *Physical review A*, 35(10):4464, 1987.

[58] Melissa Vellela and Hong Qian. Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the schlögl model revisited. *Journal of The Royal Society Interface*, 6(39):925–940, 2008.

[59] David Frigola, Laura Casanellas, José M Sancho, and Marta Ibañes. Asymmetric stochastic switching driven by intrinsic molecular noise. *PloS one*, 7(2):e31407, 2012.

[60] Robert G Endres. Bistability: Requirements on cell-volume, protein diffusion, and thermodynamics. *PloS one*, 10(4):e0121681, 2015.

[61] William H Mather, Natalie A Cookson, Jeff Hasty, Lev S Tsimring, and Ruth J Williams. Correlation resonance generated by coupled enzymatic processing. *Biophysical journal*, 99(10):3172–3181, 2010.

[62] Natalie A Cookson, William H Mather, Tal Danino, Octavio Mondragón-Palomino, Ruth J Williams, Lev S Tsimring, and Jeff Hasty. Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Molecular systems biology*, 7(1), 2011.

[63] Pablo A Iglesias and Brian P Ingalls. *Control theory and systems biology*. MIT Press, 2010.

[64] Mads Kaern, Timothy C Elston, William J Blake, and James J Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451, 2005.

[65] Donald A McQuarrie. Stochastic approach to chemical kinetics. *Journal of applied probability*, 4(3):413–478, 1967.

[66] Mukund Thattai and Alexander Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

[67] Haluk Resat, Linda Petzold, and Michel F Pettigrew. Kinetic modeling of biological systems. In *Computational Systems Biology*, pages 311–335. Springer, 2009.

[68] Tamás Székely Jr and Kevin Burrage. Stochastic simulation in systems biology. *Computational and structural biotechnology journal*, 12(20-21):14–25, 2014.

[69] Johannes Falk, Marc Mendler, and Barbara Drossel. A minimal model of burst-noise induced bistability. *PloS one*, 12(4):e0176410, 2017.

[70] Johan Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415, 2004.

[71] Ramon Grima. An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. *The Journal of chemical physics*, 133(3):07B604, 2010.

[72] Cao Li, Wu Da-Jin, and Ke Sheng-Zhi. Bistable kinetic model driven by correlated noises: unified colored-noise approximation. *Physical Review E*, 52(3):3228, 1995.

[73] Yogita Sharma and Partha Sharathi Dutta. Regime shifts driven by dynamic correlations in gene expression noise. *Physical Review E*, 96(2):022409, 2017.

[74] Tommaso Spanio, Jorge Hidalgo, and Miguel A Muñoz. Impact of environmental colored noise in single-species population dynamics. *Physical Review E*, 96(4):042301, 2017.

[75] Paul J Choi, Long Cai, Kirsten Frieda, and X Sunney Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, 322(5900):442–446, 2008.

[76] Tsz-Leung To and Narendra Maheshri. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*, 327(5969):1142–1145, 2010.

[77] Serge Pelet, Fabian Rudolf, Mariona Nadal-Ribelles, Eulàlia de Nadal, Francesc Posas, and Matthias Peter. Transient activation of the hog mapk pathway regulates bimodal gene expression. *Science*, 332(6030):732–735, 2011.

[78] Alex K Shalek, Rahul Satija, Xian Adiconis, Rona S Gertner, Jellert T Gaublomme, Raktima Raychowdhury, Schraga Schwartz, Nir Yosef,

Christine Malboeuf, Diana Lu, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, 498(7453):236, 2013.

[79] Pamela Gamba, Martijs J Jonker, and Leendert W Hamoen. A novel feedback loop that controls bimodal expression of genetic competence. *PLoS genetics*, 11(6):e1005047, 2015.

[80] Friedrich Schlögl. Chemical reaction models for non-equilibrium phase transitions. *Zeitschrift für physik*, 253(2):147–161, 1972.

[81] W Ebeling and L Schimansky-Geier. Stochastic dynamics of a bistable reaction system. *Physica A: Statistical Mechanics and its Applications*, 98(3):587–600, 1979.

[82] Chen Lihua. Systems driven by colored poisson noise: Unified colored noise approximation. *Communications in theoretical physics*, 30(1):45, 1998.

[83] Philipp Thomas, Nikola Popović, and Ramon Grima. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 111(19):6994–6999, 2014.

[84] Thomas Wilhelm. The smallest chemical reaction system with bistability. *BMC systems biology*, 3(1):90, 2009.

[85] James AJ Arpino, Edward J Hancock, James Anderson, Mauricio Barahona, Guy-Bart V Stan, Antonis Papachristodoulou, and Karen Polizzi. Tuning the dials of synthetic biology. *Microbiology*, 159(Pt 7):1236, 2013.

[86] Jeff Ross. mrna stability in mammalian cells. *Microbiol. Mol. Biol. Rev.*, 59(3):423–450, 1995.

[87] Sébastien Nouaille, Sophie Mondeil, Anne-Laure Finoux, Claire Moulis, Laurence Girbal, and Muriel Cocaign-Bousquet. The stability of an mrna is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic acids research*, 45(20):11711–11724, 2017.

[88] Ewa Grudzien-Nogalska, Joanna Kowalska, Wei Su, Andreas N Kuhn, Sergey V Slepenkov, Edward Darzynkiewicz, Ugur Sahin, Jacek Jemielity, and Robert E Rhoads. Synthetic mrnas with superior translation and stability properties. In *Synthetic messenger RNA and cell metabolism modulation*, pages 55–72. Springer, 2013.

[89] William Ashworth, Patrick N Stoney, and Tadashi Yamamoto. States of decay: The systems biology of mrna stability. *Current Opinion in Systems Biology*, 2019.

[90] William J Blake, Gábor Balázsi, Michael A Kohanski, Farren J Isaacs, Kevin F Murphy, Yina Kuang, Charles R Cantor, David R Walt, and James J Collins. Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular cell*, 24(6):853–865, 2006.

[91] Zhixing Cao and Ramon Grima. Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nature communications*, 9(1):3305, 2018.

[92] Anna Ochab-Marcinek and Marcin Tabaka. Bimodal gene expression in noncooperative regulatory systems. *Proceedings of the National Academy of Sciences*, 107(51):22096–22101, 2010.

[93] William H Mather, Jeff Hasty, Lev S Tsimring, and Ruth J Williams. Translational cross talk in gene networks. *Biophysical journal*, 104(11):2564–2572, 2013.

[94] Philipp Thomas, Hannes Matuschek, and Ramon Grima. How reliable is the linear noise approximation of gene regulatory networks? *BMC genomics*, 14(4):S5, 2013.

[95] David L Nelson, Albert L Lehninger, and Michael M Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.

[96] Desmond J Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3):525–546, 2001.

[97] Michael Assaf, Elijah Roberts, Zaida Luthey-Schulten, and Nigel Goldenfeld. Extrinsic noise driven phenotype switching in a self-regulating gene. *Physical review letters*, 111(5):058102, 2013.

[98] Giulio Caravagna, Giancarlo Mauri, and Alberto d'Onofrio. The interplay of intrinsic and extrinsic bounded noises in biomolecular networks. *PLoS One*, 8(2):e51174, 2013.

[99] Yayun Zheng, Larissa Serdukova, Jinqiao Duan, and Jürgen Kurths. Transitions in a genetic transcriptional regulatory system under lévy motion. *Scientific reports*, 6:29274, 2016.

[100] Yong Xu, Yongge Li, Hao Zhang, Xiaofan Li, and Jürgen Kurths. The switch in a genetic toggle system with lévy noise. *Scientific reports*, 6:31505, 2016.

[101] David F Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of chemical physics*, 127(21):214107, 2007.

[102] Margaritis Voliotis, Philipp Thomas, Ramon Grima, and Clive G Bowsher. Stochastic simulation of biomolecular networks in dynamic environments. *PLoS computational biology*, 12(6):e1004923, 2016.

[103] Iain G Johnston, Bernadett Gaal, Ricardo Pires das Neves, Tariq Enver, Francisco J Iborra, and Nick S Jones. Mitochondrial variability as a source of extrinsic cellular noise. *PLoS computational biology*, 8(3):e1002416, 2012.

[104] Maria Luisa Guerriero, Alexandra Pokhilko, Aurora Piñas Fernández, Karen J Halliday, Andrew J Millar, and Jane Hillston. Stochastic properties of the plant circadian clock. *Journal of The Royal Society Interface*, 9(69):744–756, 2011.

[105] Emma M Keizer, Björn Bastian, Robert W Smith, Ramon Grima, and Christian Fleck. Extending the linear-noise approximation to biochemical systems influenced by intrinsic noise and slow lognormally distributed extrinsic noise. *Physical Review E*, 99(5):052417, 2019.

[106] Marco Del Giudice, Stefano Bo, Silvia Grigolon, and Carla Bosia. On the role of extrinsic noise in microrna-mediated bimodal gene expression. *PLoS computational biology*, 14(4):e1006063, 2018.